High-throughput MD Tutorial

Aleksandra Ivanova
PhD student
Palacký University
supervisor: **Pavel Polishchuk**

RCSB PDB Deposit Search Visualize Analyze Download Learn More Documentation Careers MyPDB Contact us

198,528 Structures from the PDB
1,000,361 Computed Structure Models (CSM)

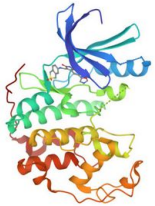
3D Structures | Enter search term(s), Entry ID(s), or sequence | Include CSM

Advanced Search | Browse Annotations | Help

PDB-101 wwPDB EMDataResource Nucleic Acid Database wwPDB Foundation

Structure Summary | 3D View | Annotations | Experiment | Sequence | Genome | Ligands | Versions

Biological Assembly 1



3D View: Structure | 1D-3D View | Electron Density | Validation Report | Ligand Interaction

Global Symmetry: Asymmetric - C1
Global Stoichiometry: Monomer - A1

Find Similar Assemblies

1KE7

Display Files | Download Files

CYCLIN-DEPENDENT KINASE 2 (CDK2) COMPLEXED WITH 3-{[(2,2-DIOXIDO-1,3-DIHYDRO-2-BENZOTHIEN-5-YL)AMINO]METHYLENE}-5-(1,3-OXAZOL-5-YL)-1,3-DIHYDRO-2H-INDOL-2-ONE

PDB DOI: 10.2210/pdb1KE7/pdb

Classification: TRANSFERASE
Organism(s): Homo sapiens
Expression System: Spodoptera frugiperda
Mutation(s): No

Deposited: 2001-11-14 **Released:** 2002-05-14
Deposition Author(s): Bramson, H.N., Corona, J., Davis, S.T., Dickerson, S.H., Edelstein, M., Frye, S.V., Gampe, R.T., Hassell, A.H., Shewchuk, L.M., Kuyper, L.F.

Experimental Data Snapshot

Method: X-RAY DIFFRACTION
Resolution: 2.00 Å
R-Value Free: 0.235
R-Value Work: 0.191

wwPDB Validation | 3D Report | Full Report

Metric	Percentile Ranks	Value
Clashscore		10
Ramachandran outliers		0
Sidechain outliers		0.4%
RSRZ outliers		4.3%

Worse ... Better

Percentile relative to all X-ray structures
Percentile relative to X-ray structures of similar resolution

Macromolecules

Find similar proteins by: Sequence (by identity cutoff) | 3D Structure

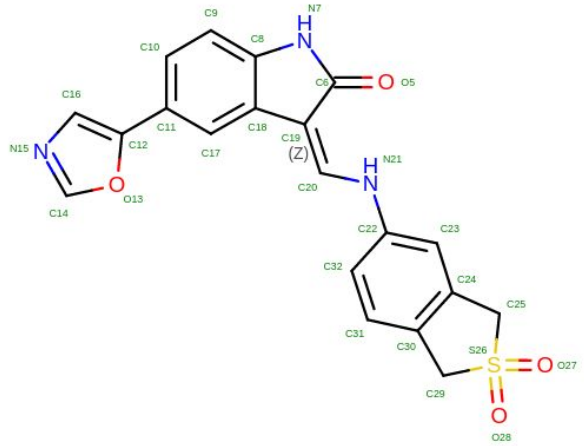
Entity ID: 1

Molecule	Chains	Sequence Length	Organism	Details	Image
Cell division protein kinase 2	A	298	Homo sapiens	**Mutation(s):** 0 **Gene Names:** CDK2, CDKN2 **EC:** 2.7.1.37 (PDB Primary Data), 2.7.11.22 (UniProt)	

UniProt & NIH Common Fund Data Resources

Find proteins for P24941 *(Homo sapiens)* Explore P24941 Go to UniProtKB: P24941

1KE7: Ligand LS3



Small Molecules

Ligands 1 Unique

ID	Chains	Name / Formula / InChI Key	2D Diagram	3D Interactions
LS3 Query on LS3 Download Ideal Coordinates CCD File Download Instance Coordinates	B [auth A]	3-{[(2,2-DIOXIDO-1,3-DIHYDRO-2-BENZOTHIEN-5-YL)AMINO]METHYLENE}-5-(1,3-OXAZOL-5-YL)-1,3-DIHYDRO-2H-INDOL-2-ONE $C_{20}H_{15}N_3O_4S$ FTQYGMLRLRXBPT-IDUWFGFVSA-N		Ligand Interaction

Binding Affinity Annotations

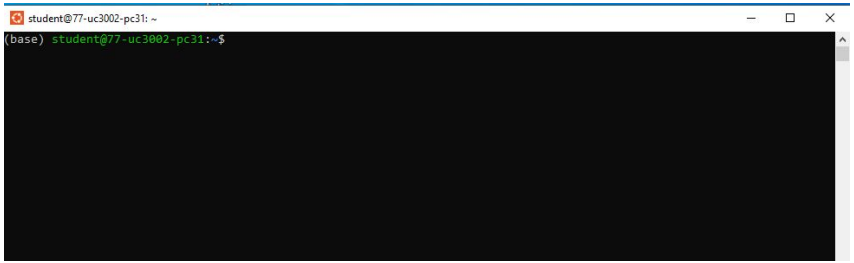
ID	Source	Binding Affinity
LS3	BindingDB: 1KE7	**IC50:** 8.9 (nM) from 1 assay(s)
	Binding MOAD: 1KE7	**IC50:** 8.9 (nM) from 1 assay(s)
	PDBBind: 1KE7	**IC50:** 8.9 (nM) from 1 assay(s)

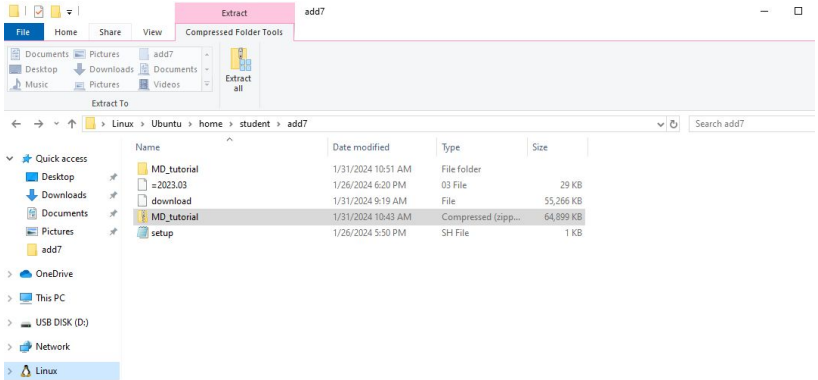
Run MD simulation (100 ps):

*download data from **https://www.kfc.upol.cz/7add***

*unzip manually under **Ubuntu/home/student/7add** directory*

cd 7add/MD_tutorial/files
conda activate md



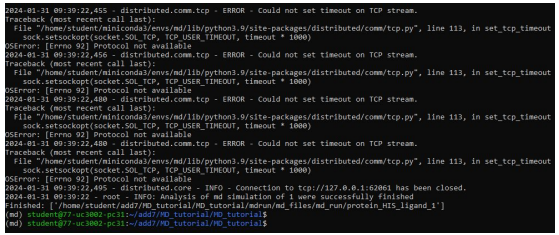


Run MD simulation (100 ps):

run_md -p protein_HIS.pdb -l ligand.mol --md_time 0.1 --nvt_time 10 --npt_time 10 --ncpu 8 -d mdrun

student@77-uc3002-pc31: ~/add7/MD_tutorial/files

```
(base) student@77-uc3002-pc31:~$ cd add7/MD_tutorial/files/
(base) student@77-uc3002-pc31:~/add7/MD_tutorial/files$ conda activate md
WARNING: No ICDs were found. Either,
- Install a conda package providing a OpenCL implementation (pocl, oclgrind, intel-compute-runtime, beignet) or
(md) student@77-uc3002-pc31:~/add7/MD_tutorial/files$ run_md -p protein_HIS.pdb -l ligand.mol --md_time 0.1 --nvt_time 10 --npt_time 10 --ncpu 8 -d mdrun
```



Linux › Ubuntu › home › student › add7 › MD_tutorial › files › mdrun

Name	Date modified	Type	Size
md_files	1/31/2024 10:54 AM	File folder	
log_protein_HIS_ligand__31-01-2024-09-2...	1/31/2024 9:39 AM	Text Document	2 KB
log_protein_HIS_ligand__31-01-2024-09-2...	1/31/2024 10:51 AM	IDENTIFIER File	0 KB
streamd_bash_protein_HIS_ligand__31-01...	1/31/2024 9:23 AM	Text Document	5 KB
streamd_bash_protein_HIS_ligand__31-01...	1/31/2024 10:51 AM	IDENTIFIER File	0 KB

log_protein_HIS_ligand__31-01-2024-09-23-54 - Notepad

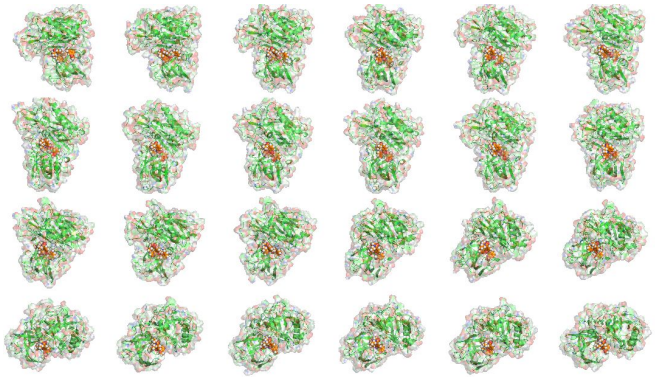
File Edit Format View Help

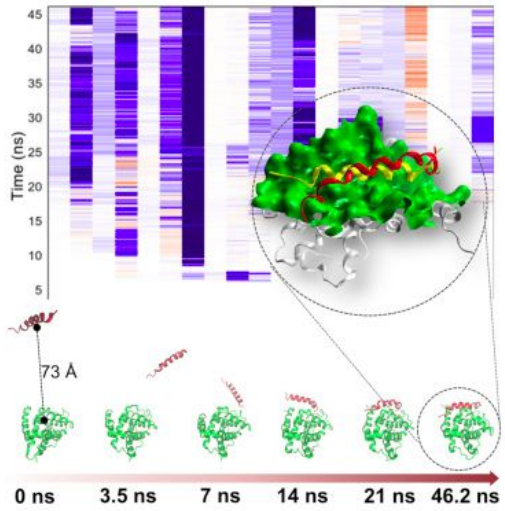
```
2024-01-31 09:23:54 - root - INFO: Namespace(protein='/home/student/add7/MD_tutorial/MD_t
2024-01-31 09:23:54 - root - INFO: Start protein preparation
2024-01-31 09:23:55 - root - INFO: Successfully finished protein preparation

2024-01-31 09:23:55 - root - INFO: Start ligand preparation
2024-01-31 09:24:34 - root - INFO: Successfully finished 1 ligand preparation

2024-01-31 09:24:36 - root - INFO: Start complex preparation
2024-01-31 09:24:38 - root - INFO: Successfully finished 1 complex preparation

2024-01-31 09:24:40 - root - INFO: Start Equilibration steps
2024-01-31 09:27:18 - root - INFO: Successfully finished 1 Equilibration step

2024-01-31 09:27:18 - root - INFO: Start Simulation step
2024-01-31 09:39:12 - root - INFO: Simulation of 1 were successfully finished
Finished: ['/home/student/add7/MD_tutorial/MD_tutorial/mdrun/md_files/md_run/protein_HIS_

2024-01-31 09:39:14 - root - INFO: Start Analysis of the simulations
2024-01-31 09:39:22 - root - INFO: Analysis of md simulation of 1 were successfully finis
Finished: ['/home/student/add7/MD_tutorial/MD_tutorial/mdrun/md_files/md_run/protein_HIS_
```

Molecular dynamics

1. MD simulations **mimic the physical motions of atoms present in the actual environment;**
2. The atoms and molecules are allowed to interact for a fixed period of time, giving **a view of the dynamic "evolution" of the system.**

- to explore conformational space
- to explore biological process of molecular recognition
- to investigate ligand pose stability
- to explore protein flexibility
- to estimate binding affinity of protein-ligand (protein-protein) complexes

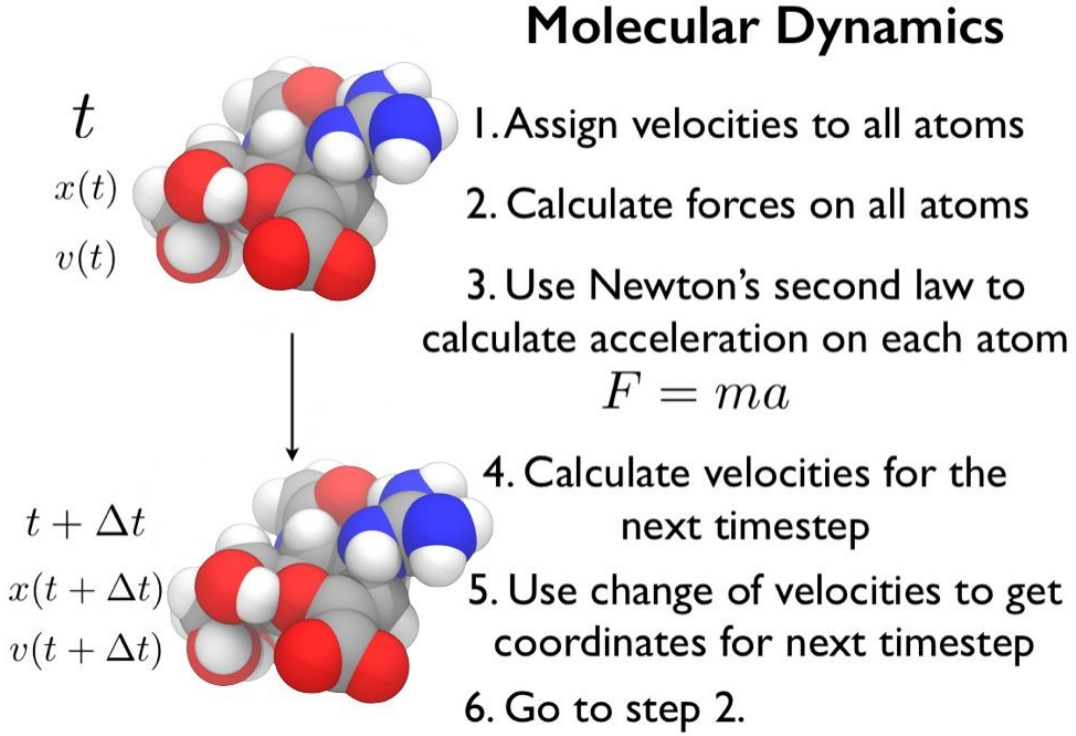




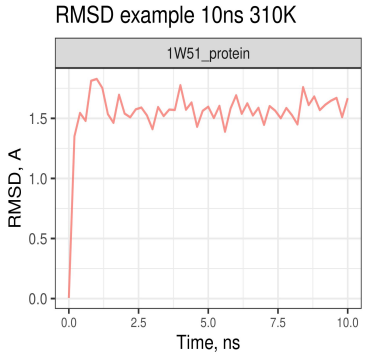
https://doi.org/10.1016/j.str.2017.02.009

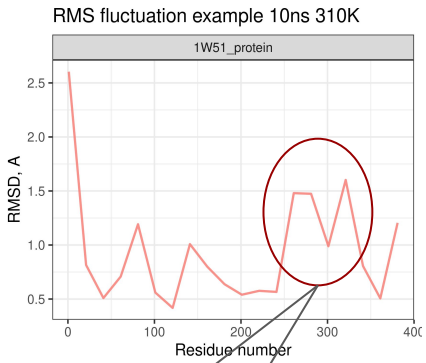
Classical Molecular Dynamics

Simulation process is based on Newton's second law

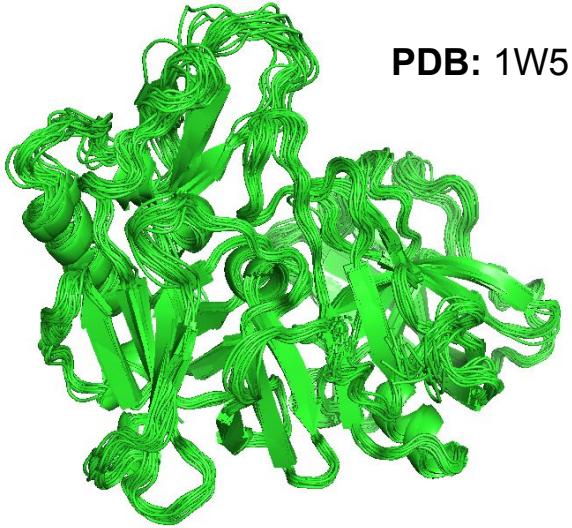


What can be done by MD

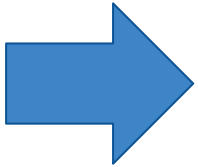




- To explore different conformation of protein
  - To investigate internal-flexibility of protein
  - For practical use we select conformations from the most populated clusters of all conformations

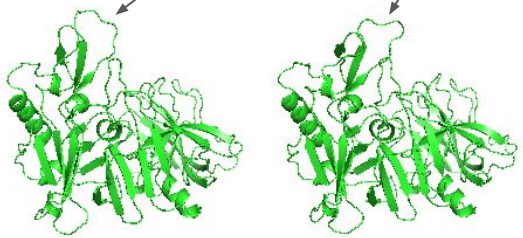


PDB: 1W51



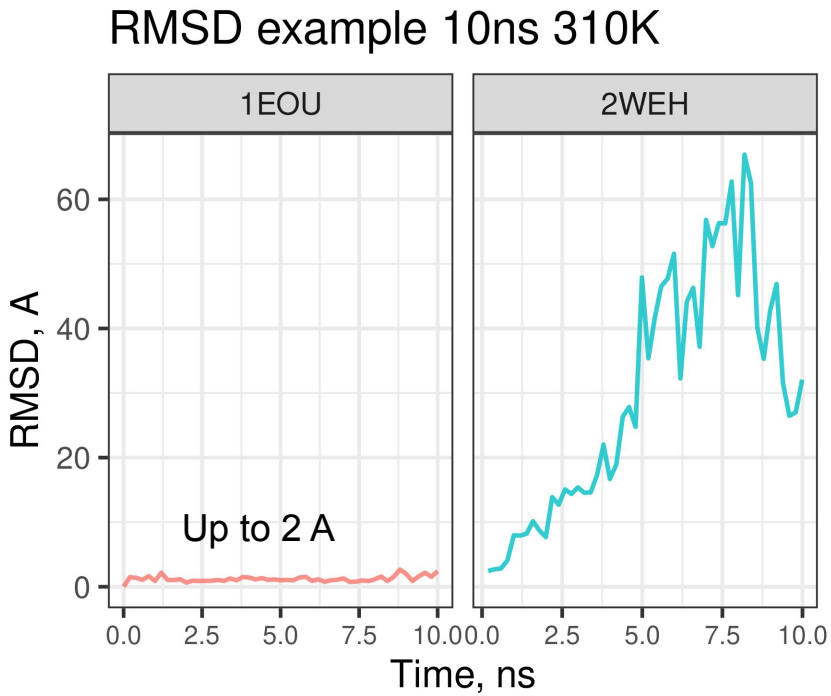
Conformation can be used:

- for further molecular docking study

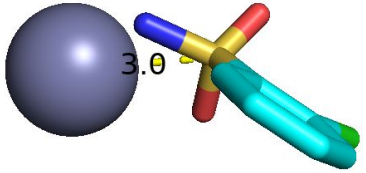


What can be done by MD

- To explore stability of ligand pose

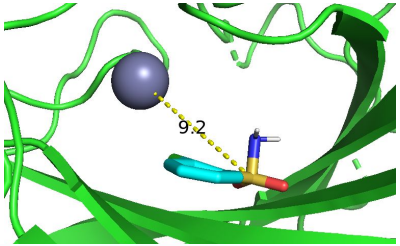


Example of incorrect protonation and pose:

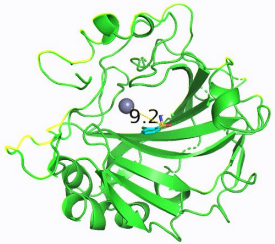


x-ray

PDB: 2WEH



docking

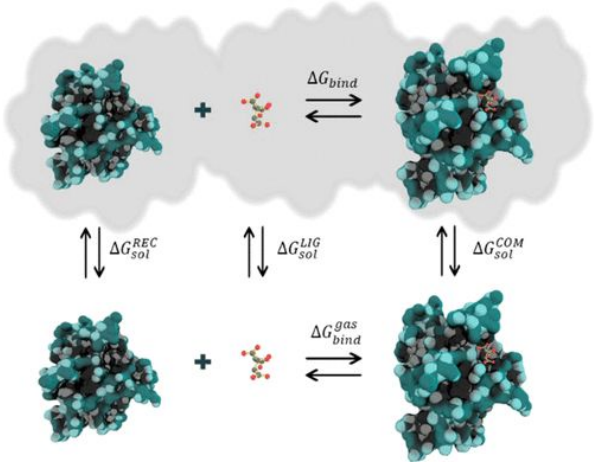


Carbonic Anhydrase

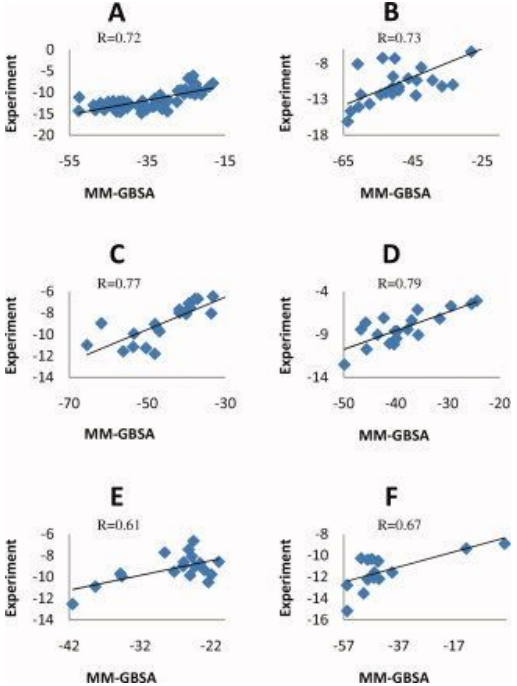
MD 10ns.
Ligand goes out of the site

What can be done by MD

- to estimate binding affinity of protein-ligand complexes



*Correlation between **MM-GBSA** predicted and experimental binding free energy.*



Yang T, Wu JC, Yan C, Wang Y, Luo R, Gonzales MB, Dalby KN, Ren P. Virtual screening using molecular simulations. Proteins. 2011 Jun;79(6):1940-51. doi: 10.1002/prot.23018. Epub 2011 Apr 12. PMID: 21491494; PMCID: PMC3092865.

Valdés-Tresanco, M.S., Valdés-Tresanco, M.E., Valiente, P.A. and Moreno E. ***gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS.*** Journal of Chemical Theory and Computation, 2021 17 (10), 6281-6291. https://pubs.acs.org/doi/10.1021/acs.jctc.1c00645.

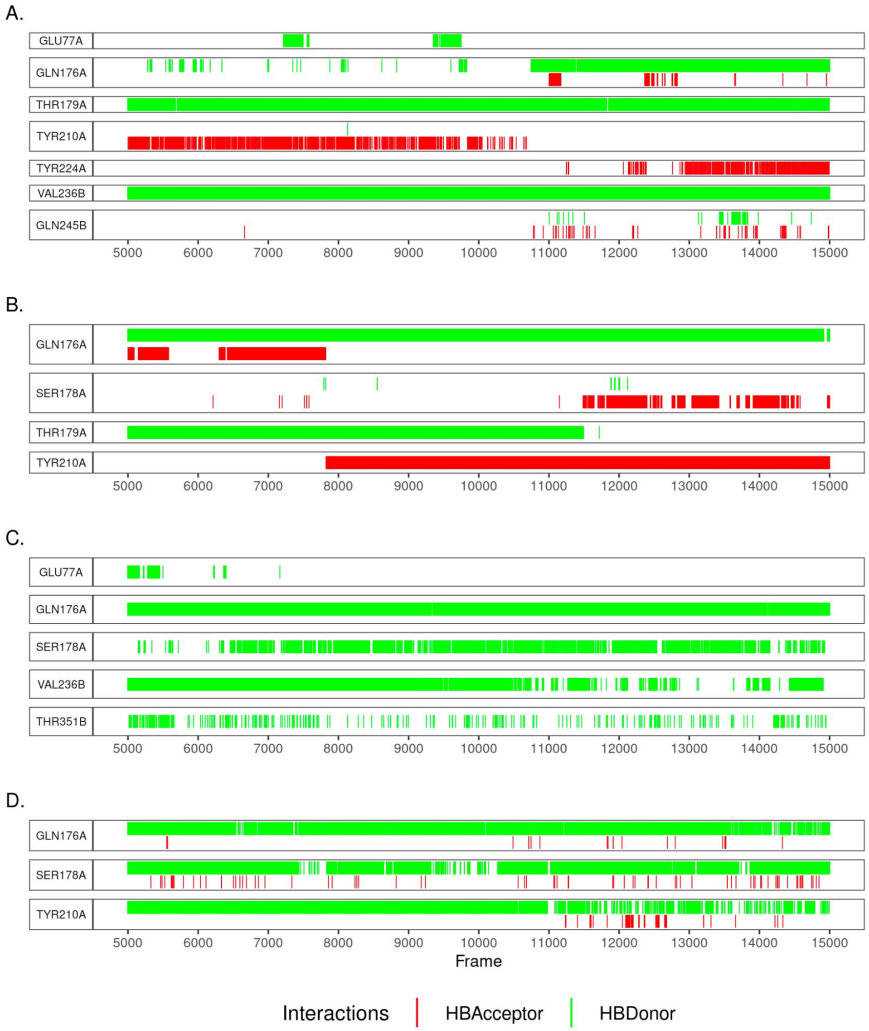
***MMPBSA.py*: An Efficient Program for End-State Free Energy Calculations** Bill R. Miller III, T. Dwight McGee Jr., Jason M. Swails, Nadine Homeyer, Holger Gohlke, and Adrian E. Roitberg Journal of Chemical Theory and Computation **2012** *8* (9), 3314-3321 DOI: 10.1021/ct300418h

What can be done by MD

- to investigate protein-ligand interaction stability

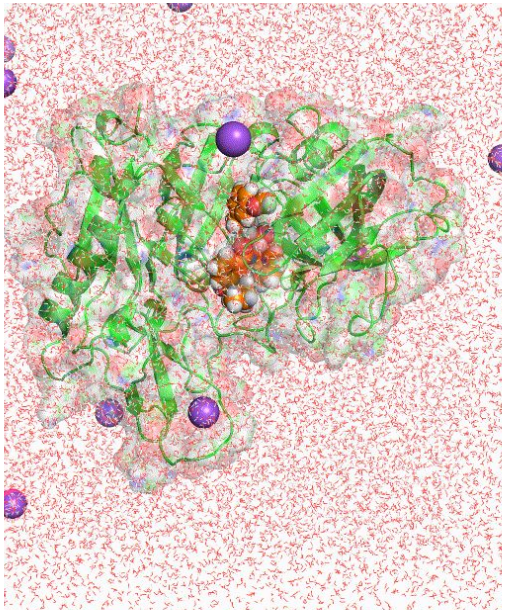
[5]:

ligand	LIG1.G						
protein	TYR38.A		TYR109.A		THR110.A	TRP125.A	
interaction	Hydrophobic	VdWContact	Hydrophobic	VdWContact	Hydrophobic	Hydrophobic	VdWContact
Frame							
0	False	False	True	False	False	True	False
10	False	False	True	True	False	True	False
20	False	False	True	True	False	True	True
30	True	False	True	False	False	True	True
40	False	False	True	False	True	True	True
50	True	False	True	True	False	False	False
60	False	False	True	False	False	False	False
70	False	False	True	True	False	True	False
80	False	False	True	False	False	True	False
90	False	False	False	False	False	True	False



Bouysset, C., Fiorucci, S. ProLIF: a library to encode molecular interactions as fingerprints.
J Cheminform 13, 72 (2021).
https://doi.org/10.1186/s13321-021-00548-6

Classical Molecular Dynamics



Structure Preparation

Preprocessing

Force-field
- Definitions of inter-atomic bonded and no-bonded forces (ligand and protein)

Simulation box setup
- box size/shape

Solvate system
- add HOH molecules

Neutralize system
- add NA+/CL- ions

Pre-simulation steps

Energy minimization
- Stop minimization when the max force < 1000.0 kJ/mol/nm
- to ensure that the system has no steric clashes or inappropriate geometry

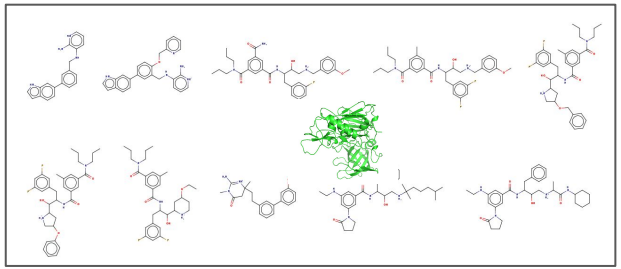
NVT and NPT equilibration
- 1000 ps
- equilibrate the solvent and ions around the protein

Product Simulation

Analysis

FAST. FLEXIBLE. FREE.
GROMACS

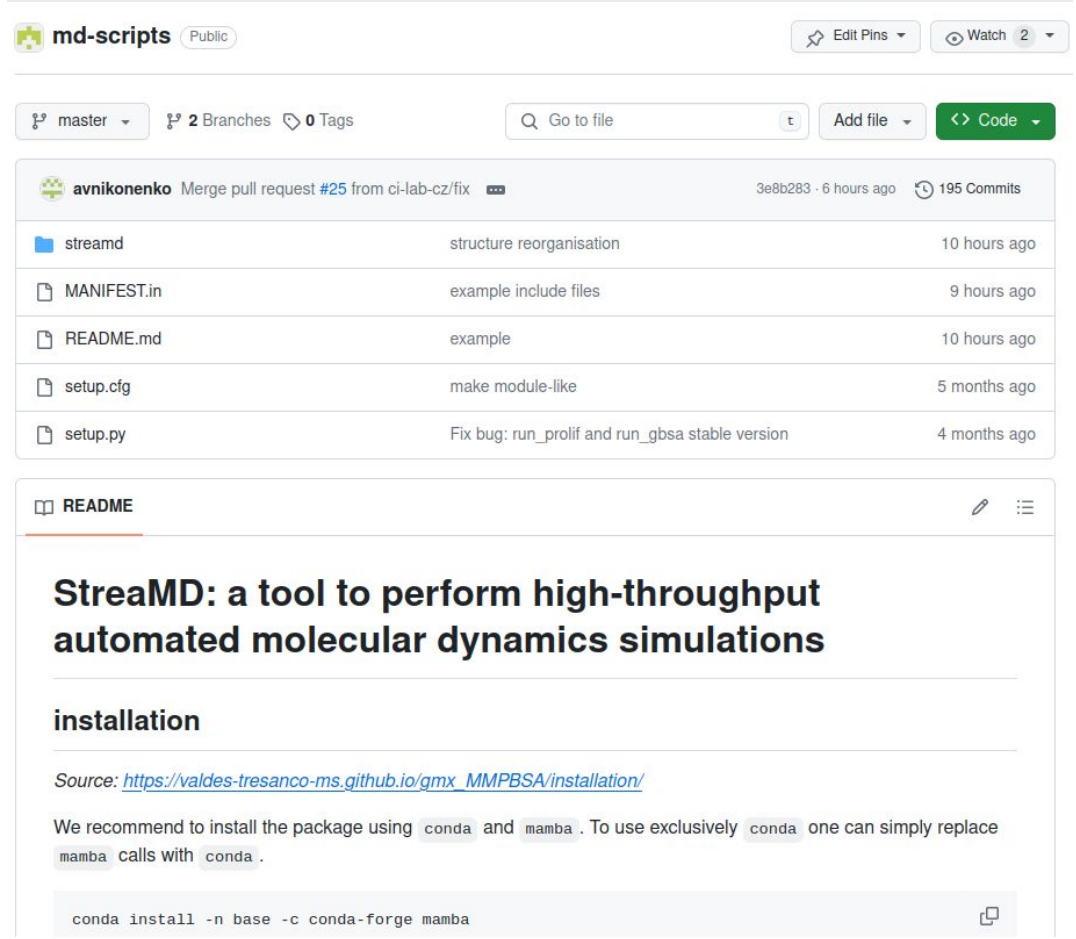
Aim: to implement an easy to use tool to run whole pipeline of Molecular Dynamics simulation automatically



Perform **High-Throughput Molecular Dynamics**

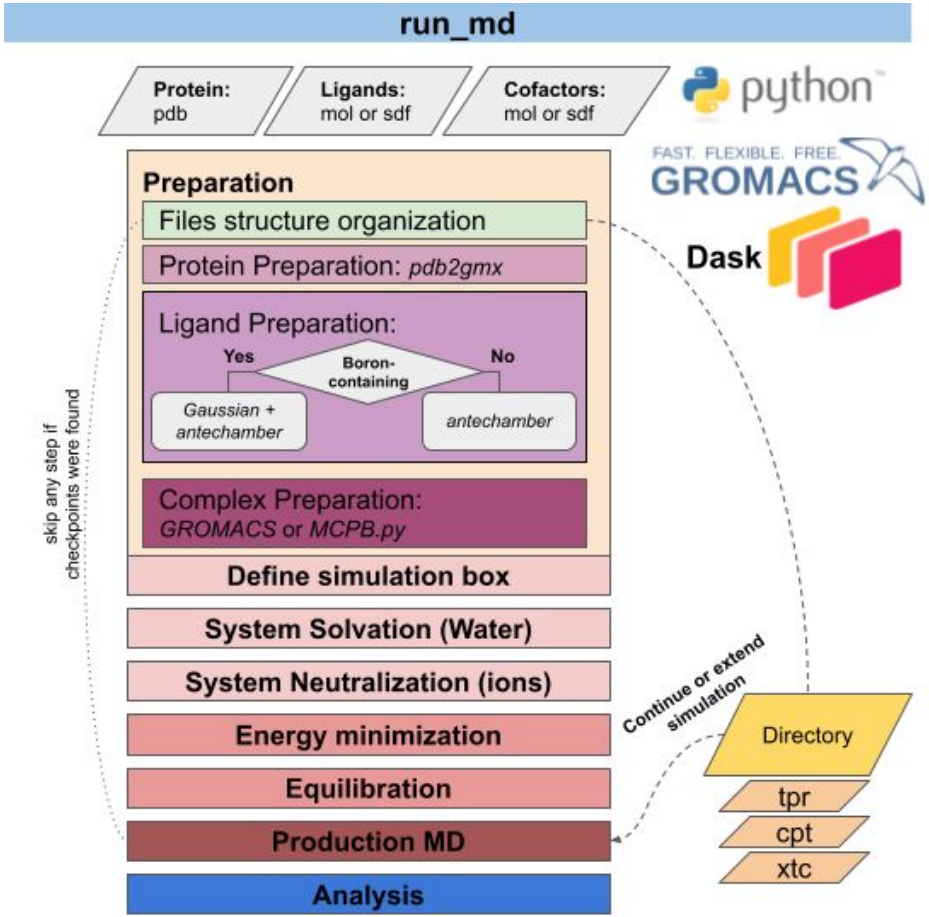
Rank compounds by:

- Stability of previously obtained pose of ligand
- Calculated Binding Energy
- Stability of protein-ligand contacts



The tool is already implemented and free available

https://github.com/ci-lab-cz/md-scripts



Automatic Analysis:

System preparation analysis:
potential.png - potential energy
temperature.png
pressure.png
density.png

Trajectory analysis:
rmsd.png - rmsd of the protein against minimized structure
rmsd_xtal.png - rmsd of the protein against crystal structure
rmsd_cofactor-molid.png - rmsd of each cofactors
rmsd_ligand-molid.png - rmsd of the ligand
rmsf.png - root mean square fluctuation (standard deviation) of atomic positions
gyrate.png- radius of gyration

run_gbsa

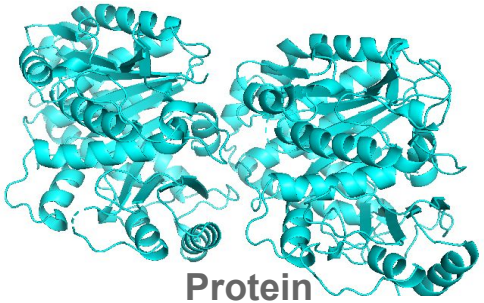
- prepare arguments and calculate MMP(G)PBSA energy using **gmx_MMPBSA** tool
- **Input:** *directory* or *top, ndx, xtc* files
- use default *mmpbsa.in* if file was not set
- automatic calculation of the appropriate MPI Processes
- parse outputs and save aggregated csv file

run_prolif

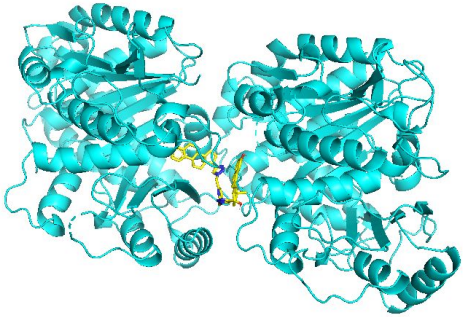
- Get protein-ligand interactions from MD trajectories using **ProLIF** module
- **Input:** directory or tpr, xtc files
- CSV and png output files

Main features of the tool:

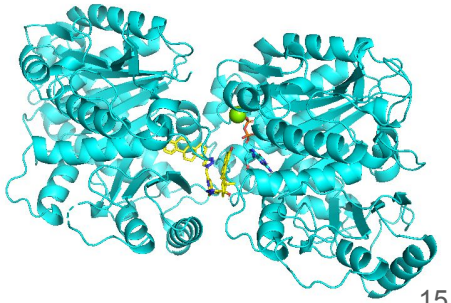
- **User control of simulation time**
  - from 10 ps to 1 μs
- **Default preset optimal parameters to run Molecular Dynamics**
  - can be easily modified
  - useful as teaching source
- **Support of modeling of different molecular systems**
  - protein only, protein-ligand, protein-ligand-cofactor(s)
- **Support of modeling of boron-containing molecules**
  - using Gaussian tool
- **Ability to continue interrupted or to expand already finished simulations**
- **Support of distributed computing using dask library provides**
- **Automatic analysis of simulation:**
  - RMSD plots for both protein and ligand objects
  - Plot of flexibility of each amino acids (RMSF)
- **Support of analysis by additional instruments:**
  - ***ProLIF***: Ligand-Protein interactions (time-dependent function, stability analysis)
  - ***MM(PB)GBSA:*** Calculation of Binding Energy



Protein

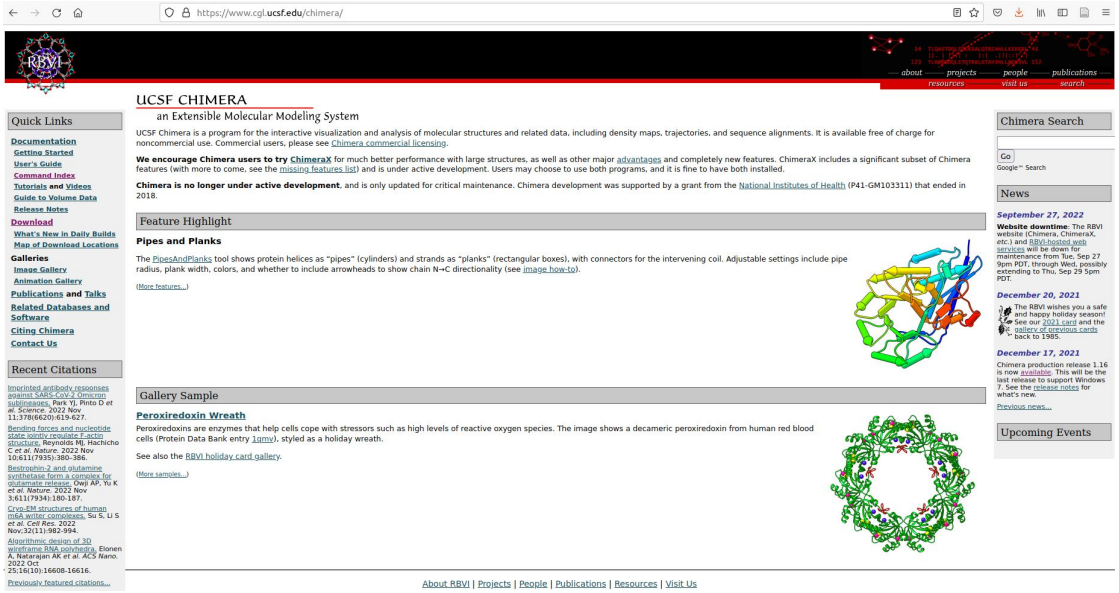


Protein-Ligand



Protein-Ligand-Cofactor

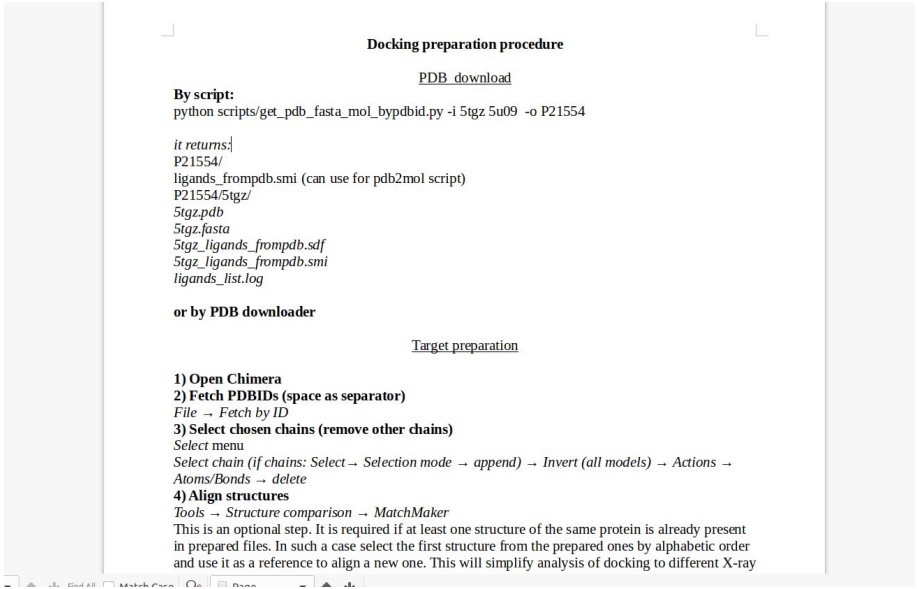
Structure preparation



https://www.cgl.ucsf.edu/chimera/

Protein and ligand preparation. Home scripts

https://github.com/ci-lab-cz/docking-files/tree/main



Docking preparation procedure

PDB_download

By script:
python scripts/get_pdb_fasta_mol_bypdbid.py -i 5tgz 5u09 -o P21554

it returns:
P21554/
ligands_frompdb.smi (can use for pdb2mol script)
P21554/5tgz/
5tgz.pdb
5tgz.fasta
5tgz_ligands_frompdb.sdf
5tgz_ligands_frompdb.smi
ligands_list.log

or by PDB downloader

Target preparation

1) Open Chimera
2) Fetch PDBIDs (space as separator)
File → Fetch by ID
3) Select chosen chains (remove other chains)
Select menu
Select chain (if chains: Select→ Selection mode → append) → Invert (all models) → Actions → Atoms/Bonds → delete
4) Align structures
Tools → Structure comparison → MatchMaker
This is an optional step. It is required if at least one structure of the same protein is already present in prepared files. In such a case select the first structure from the prepared ones by alphabetic order and use it as a reference to align a new one. This will simplify analysis of docking to different X-ray

Q9K2N0	Fix charges and Hs in mol files. Manual revision of the case…	2 months ago
Tubulin-P81947	Add a blind docking site box for tubulin	5 months ago
Tyrosinase-C7FF05	docking files	3 years ago
scripts	Fix charges and Hs in mol files. Manual revision of the case…	2 months ago
.gitignore	add Uniprot code to folder name and add mol files to ligand…	2 months ago
README.md	Update README.md	3 months ago
Target-prepare_desgn.docx	add pdbqt of boron-containing complexes	5 months ago
targets_list.csv	docking files	3 years ago

README

Complexes prepared for docking

This is a repository of protein-ligand complexes prepared for docking in PDB nad PDBQT formats.

The detailed pipeline is described in the document `Target-prepare_desgn.docx`. Please follow it if you contribute a new structure.

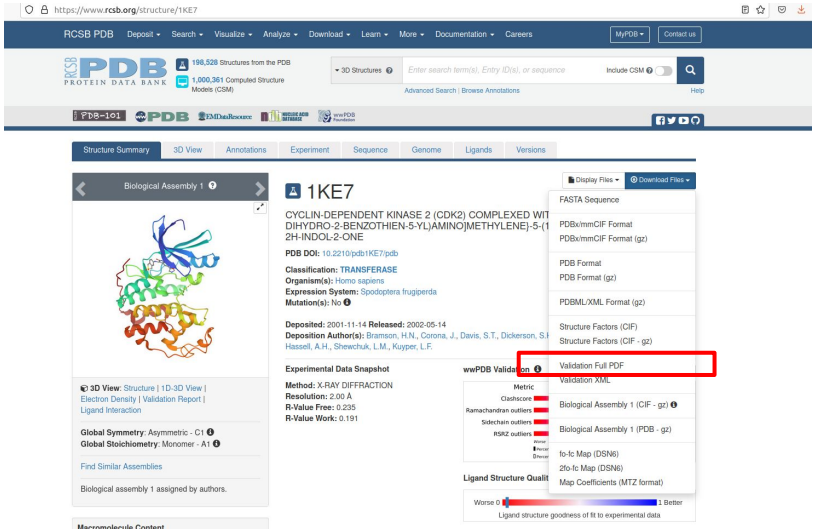
The major feature is that all structures of an individual protein are aligned and a single grid box is used for all of them.

The repository is composed of directories for individual proteins, where every directory has the following structure:

Protein preparation

1. Download structure from PDB (https://www.rcsb.org/) using PDBID

   Download Files -> PDB Format



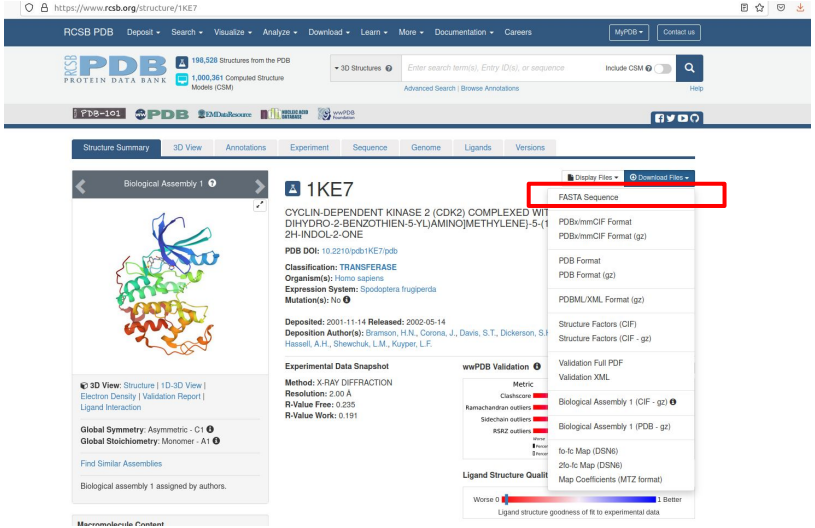
Protein preparation

1. Download structure from PDB (https://www.rcsb.org/) using PDBID

   Download Files -> PDB Format

2. Download sequence from PDB or from UniProt

   Download Files -> Fasta Sequence



Protein preparation

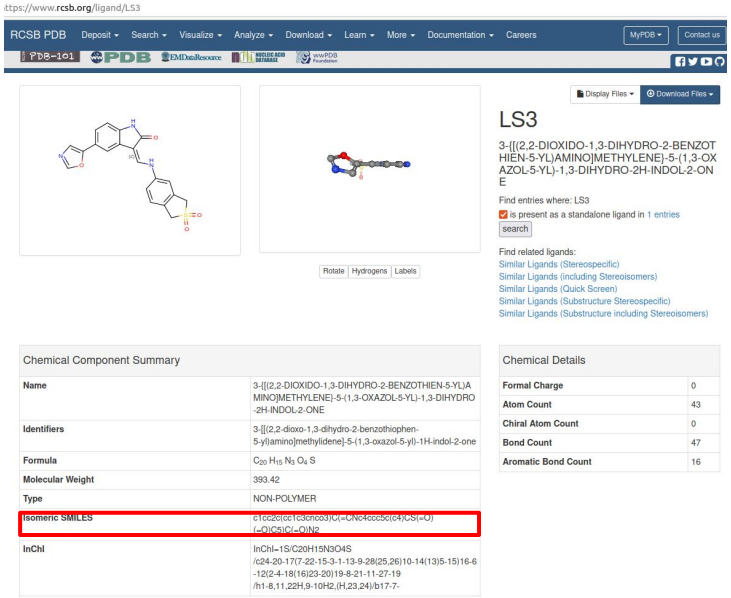
1. Download structure from PDB (https://www.rcsb.org/) using PDBID

   Download Files -> PDB Format

2. Download sequence from PDB or from UniProt

   Download Files -> Fasta Sequence

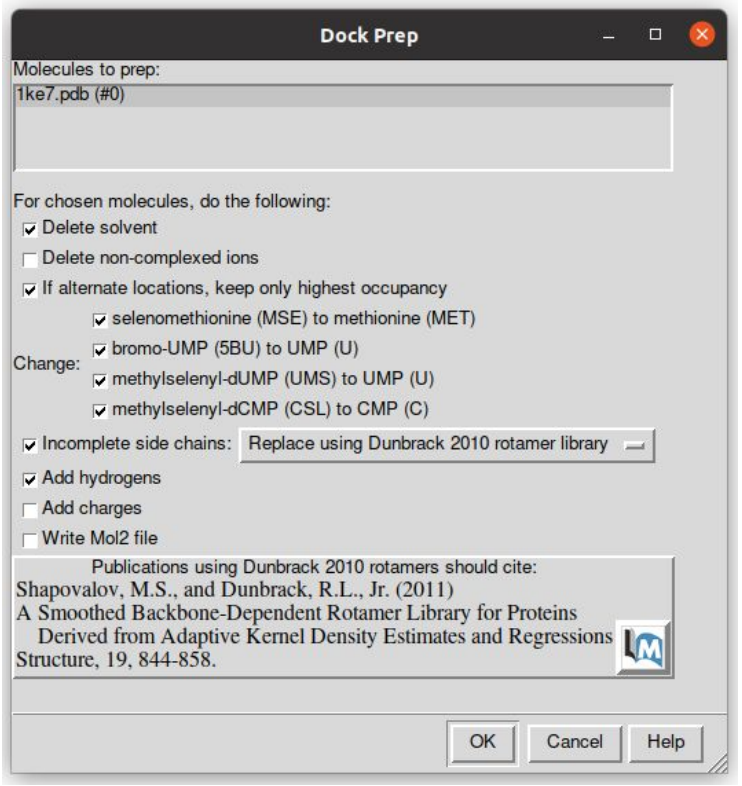
3. *Save smiles of ligand into file (ligand.smi)*



Protein preparation

4. Open Fasta and PDB in **Chimera**
    a. Dock Prepare

Structure Editing -> Dock Prep



Protein preparation

4. Open Fasta and PDB in **Chimera**

   b. fill missing loops by Modeller

      Tools -> Sequence -> Sequence

      Sequence -> Structure -> Modeller (loops/refinement)

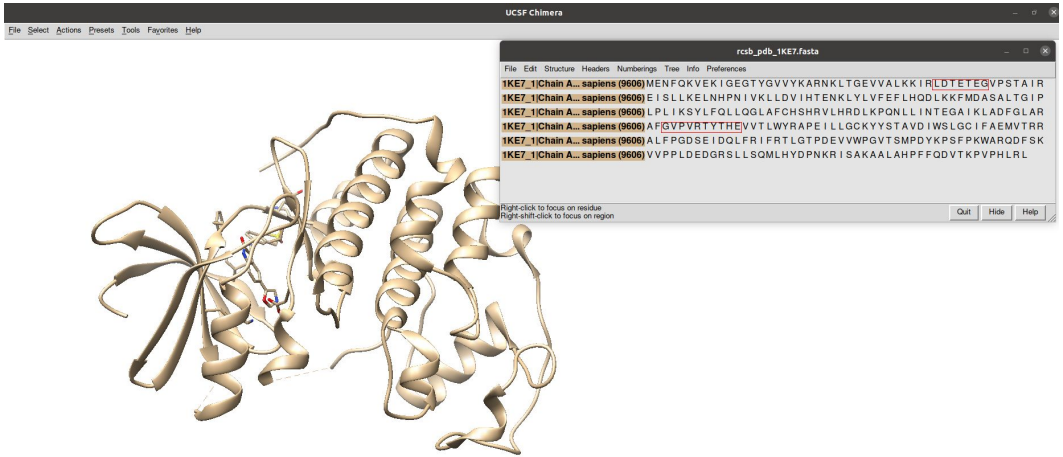
   c. *add Hs to selected model*

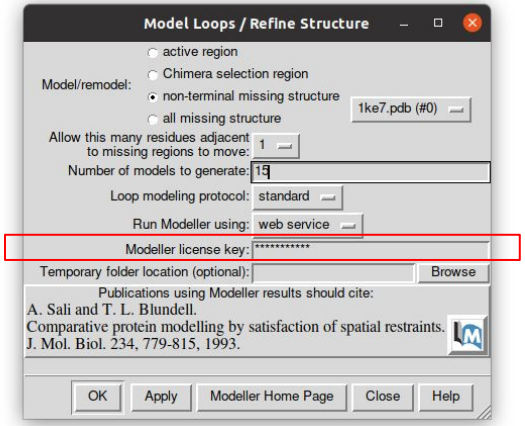
Treatment of Chosen Models

Select atoms | Choose in Model Panel | Hide others

Model	GA341	zDOPE
#1.1	1.00	-1.52
#1.2	1.00	-1.53
#1.3	1.00	-1.45
#1.4	1.00	-1.55
#1.5	1.00	-1.61
#1.6	1.00	-1.60
#1.7	1.00	-1.65
#1.8	1.00	-1.52
#1.9	1.00	-1.50
#1.10	1.00	-1.59
#1.11	1.00	-1.48
#1.12	1.00	-1.55
#1.13	1.00	-1.60
#1.14	1.00	-1.60
#1.15	1.00	-1.49

select the model with the lowest zDOPE





Protein preparation

https://ambermd.org/Questions/HIS.html

AMBER Histidine residues

Histidine (HIS in normal pdb files) is really one of three possible residues:

HID: Histidine with hydrogen on the delta nitrogen

HIE: Histidine with hydrogen on the epsilon nitrogen

HIP: Histidine with hydrogens on both nitrogens; this is positively charged.

It is up to the user to inspect the environment of each histidine and identify the type that is appropriate.

4. Open Fasta and PDB in Chimera

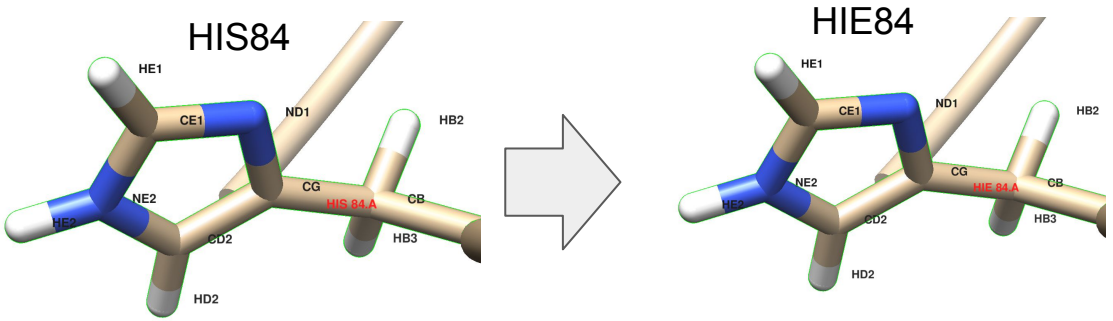
   d. Set HIS protonated state

Tools -> General Controls -> Command line

Put in the Command line:

setattr r type HID :HIS@HD1,DD1,TD1,HND
setattr r type HIP :HID@HE2,DE2,TE2
setattr r type HIE :HIS@HE2

check visually active site



Chimera preparation

4. Open Fasta and PDB in **Chimera**

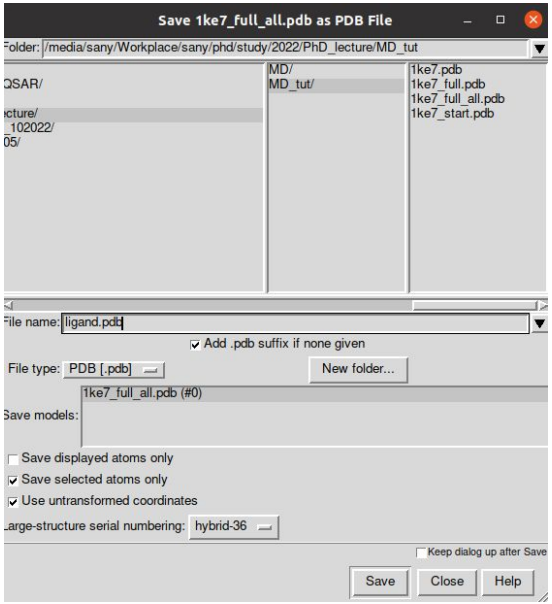
   e. Save protein only (plus ions, metals) and ligand only in *pdb* formats separately

   Select -> Residue -> Ligand_id

   File -> Save PDB -> Save selected only ligand.pdb

   Select -> Invert (all models)

   File -> Save PDB -> Save selected only protein_HIS.pdb



Protein / Ligand preparation

Input Files for MD:

protein_HIS.pdb

- *no missing non-terminal atoms/residues*
- *removed non-protein residues*
- *added all hydrogens*
- *set HID/HIE/HIP states*

ligand.mol

- *correct tautomerization*
- *protonated at 7.4 pH / 3d-based protonation (e.g. Chimera) / or user manual protonation*
- *added all hydrogens*

Ligand preparation

6. Transform pdb of ligand into mol format (add bonds information from smiles)

   Script https://github.com/ci-lab-cz/md-scripts/blob/master/scripts/pdb2mol.py

   Run in Bash/Shell:

   python md-scripts/scripts/pdb2mol.py -i ligand.pdb --smiles ligand.smi

Ligand preparation

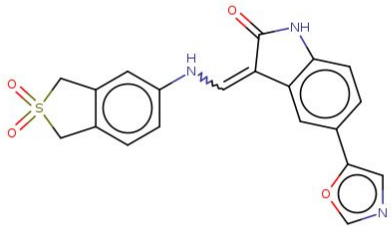
5. Save protonated smiles (pH 7.4) by Marvin into *ligand_74.smi*

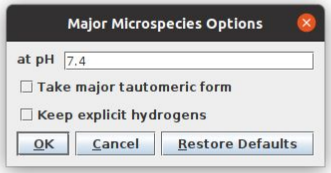
   Open Marvin Sketch

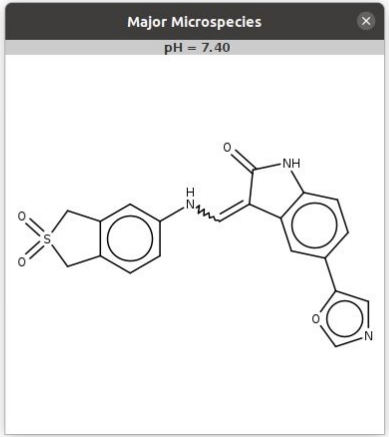
   File -> Open -> ligand.smi

   Calculations -> Protonation -> Major Microspecies -> 7.4 -> Ctrl+L (copy as Smiles)

   Save to ligand_74.smi

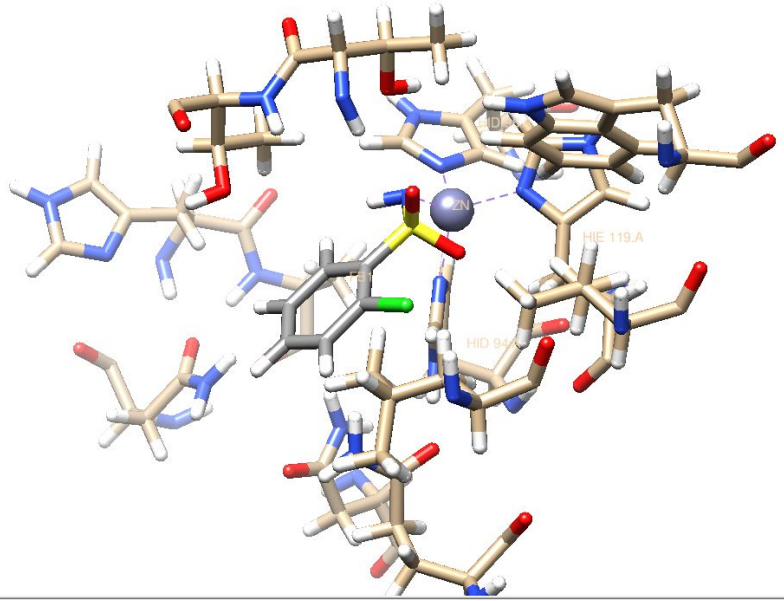


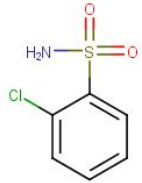


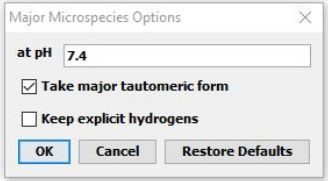


Ligand preparation

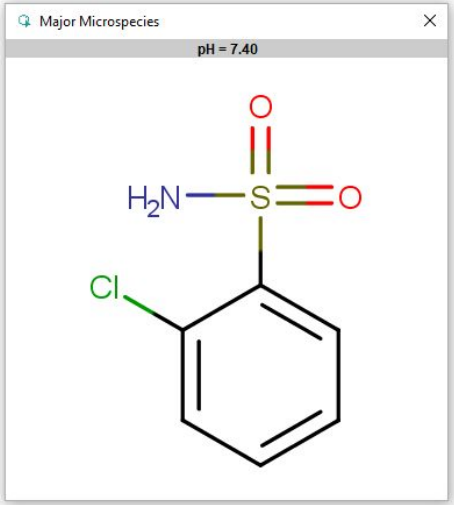
Carbonic Anhydrase







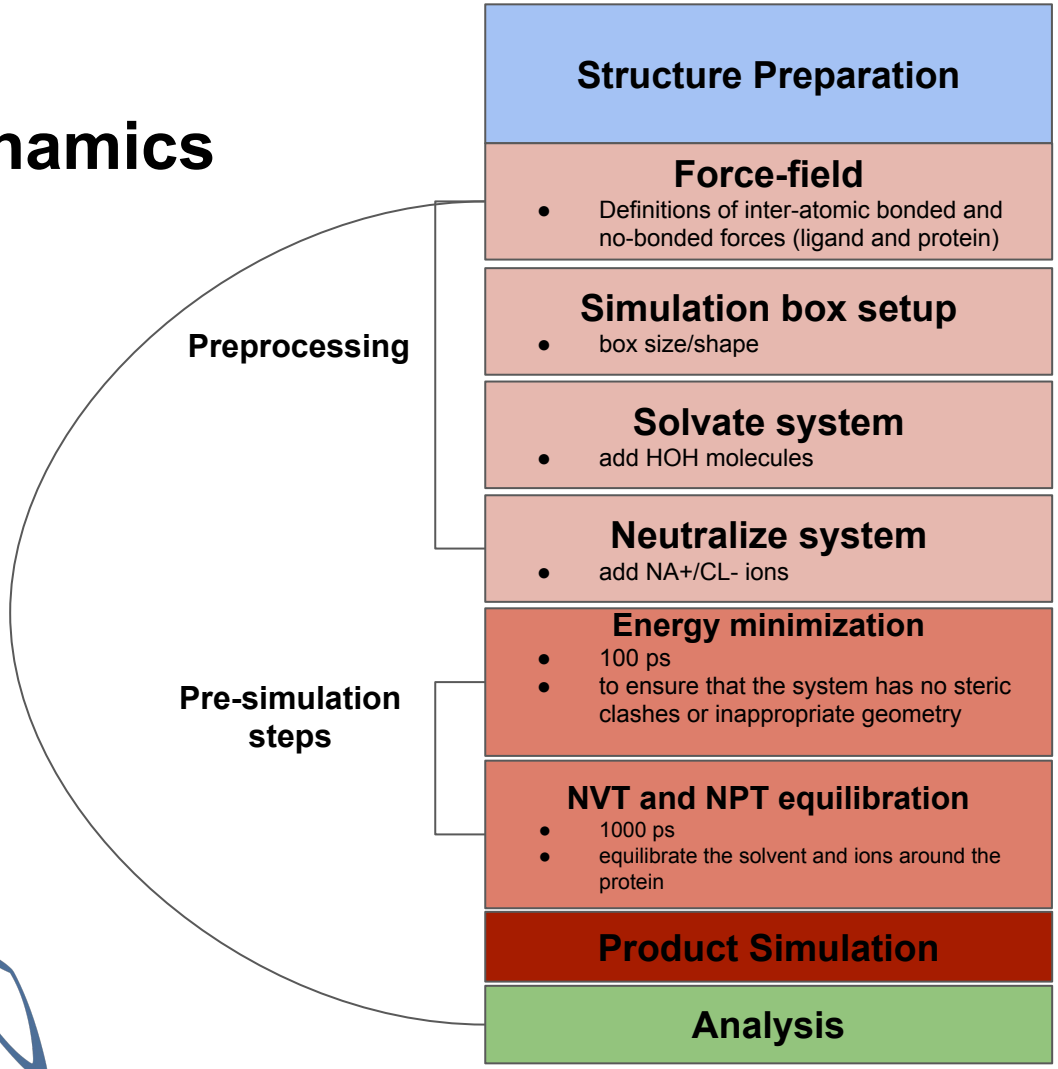
2-CHLOROBENZENESULFONAMIDE



2WEH

Classical Molecular Dynamics

https://github.com/ci-lab-cz/md-scripts



FAST. FLEXIBLE. FREE.
GROMACS

examples of *.mdp files

ions.mdp

```
; ions.mdp - used as input into grompp to generate ions.tpr
; Parameters describing what to do, when to stop and what to save
integrator  = steep         ; Algorithm (steep = steepest descent minimization)
emtol       = 1000.0        ; Stop minimization when the maximum force < 1000.0 kJ/mol/nm
emstep      = 0.01          ; Minimization step size
nsteps      = 50000         ; Maximum number of (minimization) steps to perform

; Parameters describing how to find the neighbors of each atom and how to calculate the interactions
cutoff-scheme   = Verlet    ; Buffered neighbor searching
ns_type         = grid      ; Method to determine neighbor list (simple, grid)
coulombtype     = cutoff    ; Treatment of long range electrostatic interactions
rcoulomb        = 1.0       ; Short-range electrostatic cut-off
rvdw            = 1.0       ; Short-range Van der Waals cut-off
pbc             = xyz       ; Periodic Boundary Conditions in all 3 dimensions
```

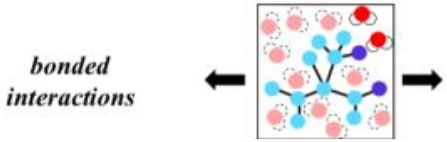
em.mdp

```
; minim.mdp - used as input into grompp to generate em.tpr
; Parameters describing what to do, when to stop and what to save
integrator  = steep         ; Algorithm (steep = steepest descent minimization)
emtol       = 1000.0        ; Stop minimization when the maximum force < 1000.0 kJ/mol/nm
emstep      = 0.01          ; Minimization step size
nsteps      = 50000         ; Maximum number of (minimization) steps to perform

; Parameters describing how to find the neighbors of each atom and how to calculate the interactions
cutoff-scheme   = Verlet    ; Buffered neighbor searching
ns_type         = grid      ; Method to determine neighbor list (simple, grid)
coulombtype     = PME       ; Treatment of long range electrostatic interactions
rcoulomb        = 1.0       ; Short-range electrostatic cut-off
rvdw            = 1.0       ; Short-range Van der Waals cut-off
pbc             = xyz       ; Periodic Boundary Conditions in all 3 dimensions
```

md.mdp

```
title                    = OPLS Lysozyme NPT equilibration
; Run parameters
integrator               = md        ; leap-frog integrator
nsteps                   = 5000000    ; 2 * 5000000 = 10000 ps (10 ns)
dt                       = 0.002     ; 2 fs
; Output control
nstxout                  = 0         ; suppress bulky .trr file by specifying
nstvout                  = 0         ; 0 for output frequency of nstxout,
nstfout                  = 0         ; nstvout, and nstfout
nstenergy                = 5000      ; save energies every 10.0 ps
nstlog                   = 5000      ; update log file every 10.0 ps
nstxout-compressed       = 5000      ; save compressed coordinates every 10.0 ps
compressed-x-grps        = System    ; save the whole system
; Bond parameters
continuation             = yes       ; Restarting after NPT
constraint_algorithm     = lincs     ; holonomic constraints
constraints              = h-bonds   ; bonds involving H are constrained
lincs_iter               = 1         ; accuracy of LINCS
lincs_order              = 4         ; also related to accuracy
; Neighborsearching
cutoff-scheme            = Verlet    ; Buffered neighbor searching
ns_type                  = grid      ; search neighboring grid cells
nstlist                  = 10        ; 20 fs, largely irrelevant with Verlet scheme
rcoulomb                 = 1.0       ; short-range electrostatic cutoff (in nm)
rvdw                     = 1.0       ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype              = PME       ; Particle Mesh Ewald for long-range electrostatics
pme_order                = 4         ; cubic interpolation
fourierspacing           = 0.16      ; grid spacing for FFT
; Temperature coupling is on
tcoupl                   = V-rescale             ; modified Berendsen thermostat
tc-grps                  = Protein Non-Protein   ; two coupling groups - more accurate
tau_t                    = 0.1     0.1           ; time constant, in ps
ref_t                    = 300     300           ; reference temperature, one for each group, in K
; Pressure coupling is on
pcoupl                   = Parrinello-Rahman     ; Pressure coupling on in NPT
pcoupltype               = isotropic             ; uniform scaling of box vectors
tau_p                    = 2.0                   ; time constant, in ps
ref_p                    = 1.0                   ; reference pressure, in bar
compressibility          = 4.5e-5                ; isothermal compressibility of water, bar^-1
; Periodic boundary conditions
pbc                      = xyz       ; 3-D PBC
; Dispersion correction
DispCorr                 = EnerPres  ; account for cut-off vdW scheme
; Velocity generation
gen_vel                  = no        ; Velocity generation is off
```

Theory: molecular interactions



In molecular dynamics (MD), force fields are mathematical models that describe the potential energy of a molecular system as a function of the spatial coordinates of its atoms.

Bond stretching

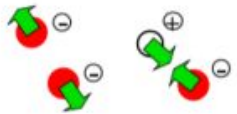
$$V^{bond}(\mathbf{r}^N) = \sum_{bonds\ i} \frac{1}{2} K_i^b \left[b_i(\mathbf{r}^N) - b_i^0 \right]^2$$

Angle bending

$$V^{angle}(\mathbf{r}^N) = \sum_{angles\ i} \frac{1}{2} K_i^\theta \left[\theta_i(\mathbf{r}^N) - \theta_i^0 \right]^2$$

Rotation around bond

$$V^{torsion}(\mathbf{r}^N) = \sum_{torsion\ i} K_i^\varphi \cos\left[m_i \varphi_i(\mathbf{r}^N) + \delta_i \right]$$



Electrostatic interactions

van der Waals interactions

All atoms see each other through space

~ ½ N (N – 1) interactions

The computational bottleneck

The force fields **define the interactions between atoms and molecules, providing a representation of the forces acting on particles within a simulation.**

Theory: force fields

All force fields were developed using different experimental data and for slightly different purposes

The most popular FFs for protein and protein-ligand simulations are:

CHARMM – proteins, lipids, nucleic acids

AMBER – peptide, protein, nucleic acids, lipid14, GAFF

OPLS - optimized to fit experimental properties of liquids, such as density and heat of vaporization

GROMOS – proteins, small molecules, also optimized for experimental parameters

Each FF works with a particular set of simulation parameters therefore not all of them can be mixed

Practice: force fields

Prepare the protein topology with pdb2gmx

gmx pdb2gmx

Synopsis

```
gmx pdb2gmx [-f [<.gro/.g96/...>]] [-o [<.gro/.g96/...>]] [-p [<.top>]]
            [-i [<.itp>]] [-n [<.ndx>]] [-q [<.gro/.g96/...>]]
            [-chainsep <enum>] [-merge <enum>] [-ff <string>]
            [-water <enum>] [-[no]inter] [-[no]ss] [-[no]ter]
            [-[no]lys] [-[no]arg] [-[no]asp] [-[no]glu] [-[no]gln]
            [-[no]his] [-angle <real>] [-dist <real>] [-[no]una]
            [-[no]ignh] [-[no]missing] [-[no]v] [-posrefc <real>]
            [-vsite <enum>] [-[no]heavyh] [-[no]deuterate]
            [-[no]chargegrp] [-[no]cmap] [-[no]renum] [-[no]rtpres]
```

Description

`gmx pdb2gmx` reads a .pdb (or .gro) file, reads some database files, adds hydrogens to the molecules and generates coordinates in GROMACS (GROMOS), or optionally .pdb, format and a topology in GROMACS format. These files can subsequently be processed to generate a run input file.

New files:
protein.gro: a molecular structure in Gromos87 format.
topol.top: molecular topology
posre.itp: position restraints

gmx pdb2gmx -f protein.pdb -o protein.gro -water tip3p -ignh <<< “AMBER99SB-ILDN”

https://manual.gromacs.org/documentation/current/onlinehelp/gmx-pdb2gmx.html

Practice: force fields

Prepare the protein topology with pdb2gmx

gmx pdb2gmx

Synopsis

```
gmx pdb2gmx [-f [<.gro/.g96/...>]] [-o [<.gro/.g96/...>]] [-p [<.top>]]
            [-i [<.itp>]] [-n [<.ndx>]] [-q [<.gro/.g96/...>]]
            [-chainsep <enum>] [-merge <enum>] [-ff <string>]
            [-water <enum>] [-[no]inter] [-[no]ss] [-[no]ter]
            [-[no]lys] [-[no]arg] [-[no]asp] [-[no]glu] [-[no]gln]
            [-[no]his] [-angle <real>] [-dist <real>] [-[no]una]
            [-[no]ignh] [-[no]missing] [-[no]v] [-posrefc <real>]
            [-vsite <enum>] [-[no]heavyh] [-[no]deuterate]
            [-[no]chargegrp] [-[no]cmap] [-[no]renum] [-[no]rtpres]
```

pdb2gmx does not work on ligand

Description

`gmx pdb2gmx` reads a .pdb (or .gro) file, reads some database files, adds hydrogens to the molecules and generates coordinates in GROMACS (GROMOS), or optionally .pdb, format and a topology in GROMACS format. These files can subsequently be processed to generate a run input file.

gmx pdb2gmx -f protein.pdb -o protein.gro -water tip3p -ignh <<< “AMBER99SB-ILDN”

https://manual.gromacs.org/documentation/current/onlinehelp/gmx-pdb2gmx.html

Automation tools for ligand topology

AMBER	**Antechamber**	Parametrizes molecules using GAFF
	acpype	A Python interface to Antechamber, writes GROMACS topologies
CHARMM	**CGenFF**	The official CHARMM General Force Field server
GROMOS87/ GROMOS96	**PRODRG 2.5**	An automated server for topology generation
	ATB	A newer server for topology generation, uses GROMOS96 54A7
OPLS-AA	**Topolbuild**	Converts a Tripos .mol2 file into a topology
	TopolGen	A Perl script to convert an all-atom .pdb file to a topology
	LigParGen	A server from the Jorgensen group to produce OPLS topologies

Ligand Force fields

Prepare the ligand topology using external tools

```
antechamber -i ligand.pdb -fi pdb -o lig.mol2 -fo mol2 -c bcc -pf y -s 2
parmchk2 -i lig.mol2 -f mol2 -o lig.frcmod
python pmed_amb2gmx.py -p lig.prmtop -x lig.inpcrd -o UNL
```

Build the topology for the protein-ligand system

Insert a line that says #include lig.itp" into topol.top after the position restraint file is included

Make adjustments is in the [molecules] directive

New files:

lig.mol2 *- file of ligand with charges and bonds*
lig.frcmod *- Parameter modification file*
posre_ligand.itp *- positional restraints*
ligand.gro *- ligand in Gromos87 format*
ligand.inpcrd - *Amber coordinate file of ligand*
ligand.prmtop *- Amber topology file*
ligand.top *- Gromacs topology file*

Changed files: topol.top

Solvation

To perform simulation we should add water molecules to mimic real environment.

There are two steps to defining the box and filling it with solvent:

1) Define the box dimensions using the editconf module
2) Fill the box with water using the solvate module.

We will use a simple cubic box as the unit cell.
First define the box using editconf:

```
gmx editconf -f protein.gro -o protein_newbox.gro -c -d 1.0 -bt cubic
```

a solute-box distance of 1.0 nm will mean that there are at least 2.0 nm between any two periodic images of a protein. This distance should be sufficient for just about any cutoff scheme commonly used in simulations.

New files:
protein_newbox.gro: protein+box system

Solvation

To perform simulation we should add water molecules to mimic real environment.

On the previous step we have defined a box, then we should fill it with solvent (water). Solvation is accomplished using solvate command:

gmx solvate -cp protein_newbox.gro -cs spc216.gro -o protein_solv.gro -p topol.top

-cp - the configuration of the protein (-cp)

-cs - the configuration of the solvent (-cs)

spc216.gro is a generic equilibrated 3-point solvent model (suitable for SPC, SPC/E, or TIP3P water, since they are all three-point water models)

New files:
protein_solv.gro: *protein + solvent system*
Changed files:
topol.top

Adding ions

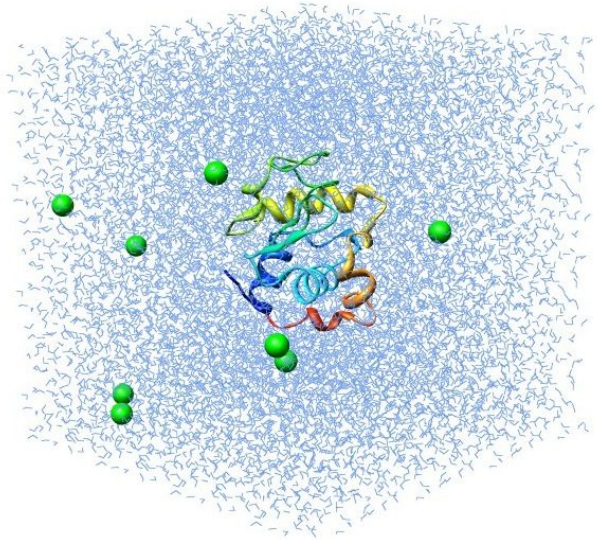
We need to add ions of CL- or NA+ to neutralize our protein:

```
gmx grompp -f ions.mdp -c protein_solv.gro -p topol.top -o ions.tpr
```

```
gmx genion -s ions.tpr -o protein_solv_ions.gro -p topol.top -pname NA -nname CL -neutral
```

New files:
protein_solv_ions.gro: *protein + solvent + ions system*
ions.tpr: *portable binary run input file. This file contains the starting structure, the molecular topology and all the simulation parameters.*

Changed files: topol.top



Solvated protein with ions

Energy minimization

Before we can begin dynamics, we must ensure that the system has no steric clashes or inappropriate geometry (equilibrium state). **The structure is relaxed through a process called energy minimization (EM)**.

Prepare files:

```
gmx grompp -f minim.mdp -c protein_solv_ions.gro -p topol.top -o  em.tpr
```

Run the minimization:

```
gmx mdrun -v -deffnm em –s em.tpr
```

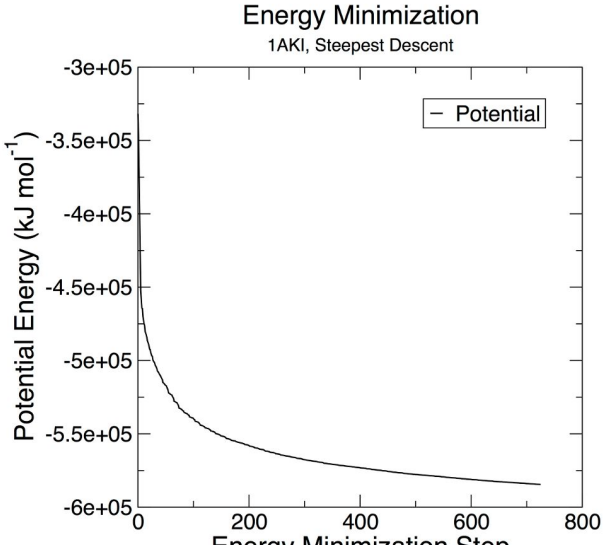
New files:
em.log: ASCII-text log file of the EM process
em.edr: Binary energy file
em.tpr: Contains the starting structure, the molecular topology and all the simulation parameters.
em.trr: Binary full-precision trajectory
em.gro: Energy-minimized structure

Energy minimization

How to do an analysis:

The em.edr file contains all of the energy terms that GROMACS collects during EM. You can analyze any .edr file using the GROMACS energy module:

gmx energy -f em.edr -o potential.xvg << "Potential"



- **Epot** should be negative, and (for a simple protein in water) on the order of 10^5-10^6,
- **maximum force**, **Fmax**, "emtol = 1000.0" should be no greater than target 1000 kJ mol-1 nm-1.

Equilibration

Energy Minimization ensured that we have a reasonable starting structure, but to begin real dynamics, we must equilibrate the solvent and ions around the protein.

Why do we need equilibration?

If we were to attempt unrestrained dynamics at this point, the system may collapse.

The reason is that the solvent is mostly optimized within itself, and not necessarily with the solute. It needs to be brought to the temperature we wish to simulate and establish the proper orientation around the solute (the protein). After we achieved the correct temperature (based on kinetic energies), we will apply pressure to the system until it reaches the proper density.

Controlling the system

Thermodynamic system has a number of state variables which describe its macroscopic state such as

- Particle number, volume, temperature, pressure, total energy

They are not all independent, but connected by equations of state

Example: Ideal gas of non-interacting point particles

$p = p(V,T)$
$E = E(V,T)$

$p \cdot V = N \cdot k_B \cdot T$
$p = \rho k_B T$
$E = \frac{3}{2} N k_B T$

https://web.mst.edu/~vojtat/class_5403/MolecularDynamics.pdf

Restrain the system

To equilibrate the solvent and ions around the protein we need to apply position restraints on the heavy atoms of the protein.

The purpose of **posre.itp** is to apply a position restraining force on the heavy atoms of the protein (anything that is not a hydrogen). Movement is permitted, but only after overcoming a substantial energy penalty.

NVT equilibration

Canonical ensemble (NVT)

- Particle number N
- Volume V
- Temperature T

} External parameters

- Total energy E
- Pressure P

} Observables to be calculated

In MD simulation: some state variables are external parameters, others are observables to be calculated

Requires a **thermostat**, an algorithm that adds and removes energy to keep the temperature constant

- Velocity rescaling based on equipartition theorem

$$\langle E_{kin} \rangle = \frac{1}{2} \sum_i m_i \langle \vec{v}_i^{\,2} \rangle = \frac{3}{2} N k_B T$$

- Berendsen thermostat, Anderson thermostat

https://web.mst.edu/~vojtat/class_5403/MolecularDynamics.pdf

NVT equilibration

to run NVT equlibration:

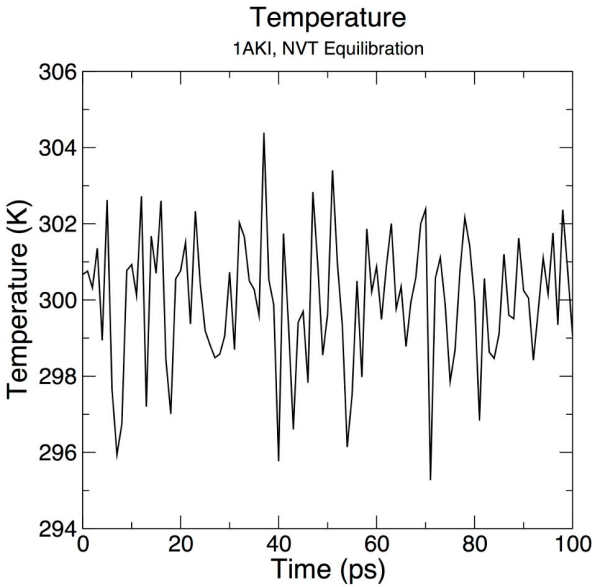
gmx grompp -f nvt.mdp -c em.gro -r em.gro -p topol.top -o nvt.tpr

gmx mdrun -deffnm nvt -s nvt.tpr

An analysis:

gmx energy -f nvt.edr -o temperature.xvg <<< “Temperature”

New files:
nvt.log: *ASCII-text log file of the equilibration process*
nvt.edr: *Binary energy file*
nvt.trr: *Binary full-precision trajectory*
nvt.gro: *NVT-minimized structure*



NPT equilibration

Isothermal–isobaric ensemble (NPT)

- Particle number N
- Pressure P
- Temperature T

} External parameters

- Total energy E
- Volume V

} Observables to be calculated

Requires a **barostat** in addition to the thermostat, an algorithm that changes volume to keep the pressure constant

https://web.mst.edu/~vojtat/class_5403/MolecularDynamics.pdf

NPT equilibration

Note that we are now including the -t flag to include the checkpoint file from the NVT equilibration

The coordinate file (-c) is the final output of the NVT simulation.

to run NPT equlibration:

gmx grompp -f npt.mdp -c nvt.gro -r nvt.gro -t nvt.cpt -p topol.top -o npt.tpr gmx mdrun -deffnm npt -s npt.tpr

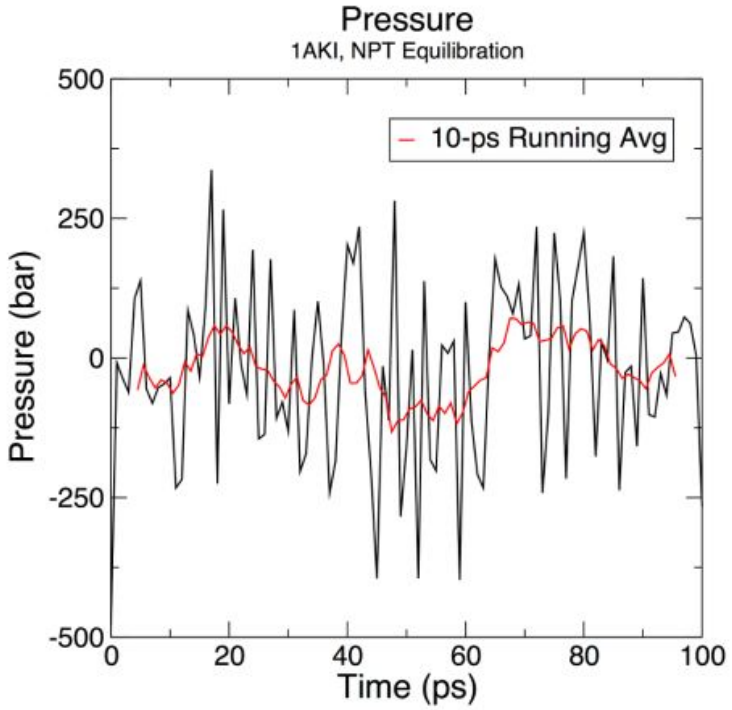
An analysis:

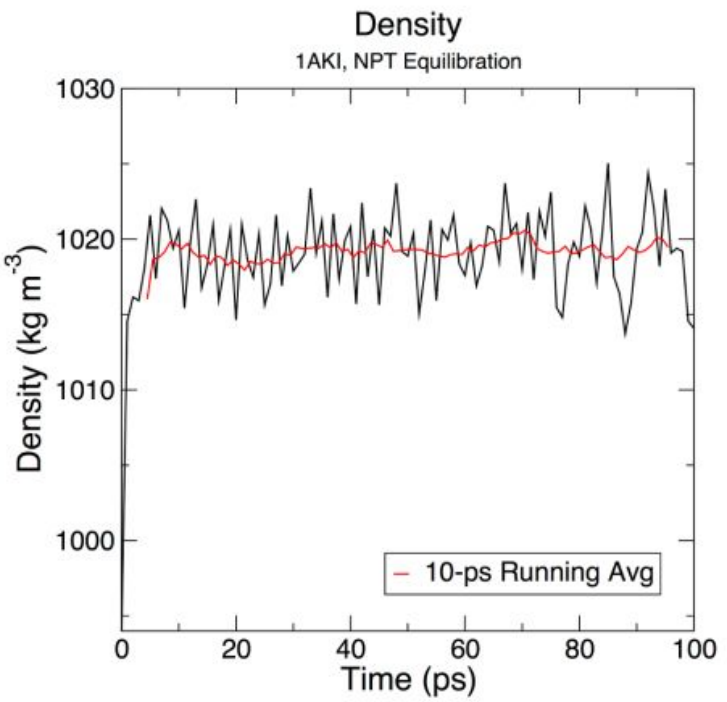
gmx energy -f npt.edr -o pressure.xvg <<< ”Pressure”

gmx energy -f npt.edr -o density.xvg <<< “Density”

New files:
nvt.log: *ASCII-text log file of the equilibration process*
nvt.edr: *Binary energy file*
nvt.trr: *Binary full-precision trajectory*
nvt.gro: *NVT-minimized structure*

NPT equilibration

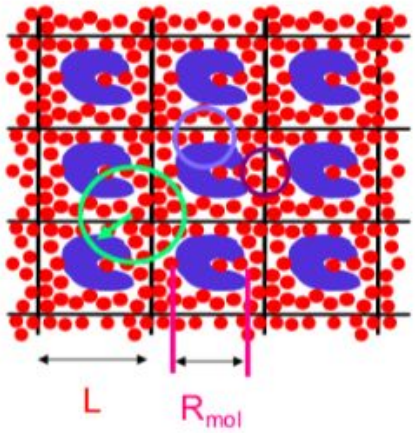




Periodic Boundary Conditions

To simulate an infinite system, **periodic boundary conditions** are often applied. This allows atoms that leave one side of the simulation box to re-enter on the opposite side.

- PBCs are often used in computer simulations and mathematical models.
- The topology of two-dimensional PBC is equal to that of a world map of some video games; the geometry of the unit cell satisfies perfect two-dimensional tiling, and when an object passes through one side of the unit cell, it reappears on the opposite side with the same velocity.



Required
(no atom sees another one twice):

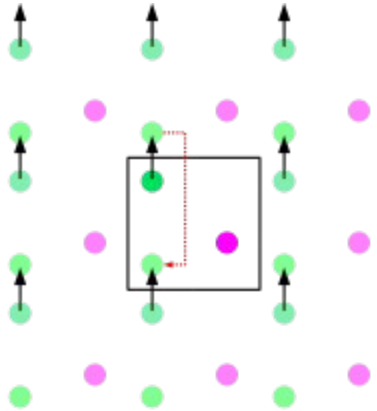
$L > 2R_c$ $R_c < ½L$

Preferred
(protein does not see a copy of itself)

$L > R_c + R_{mol}$ $R_c < L - R_{mol}$

Even better
(no solvent sees two proteins)

$L > 2R_c + R_{mol}$ $R_c < ½(L - R_{mol})$



- Cubic:

Rectangular:

Production MD

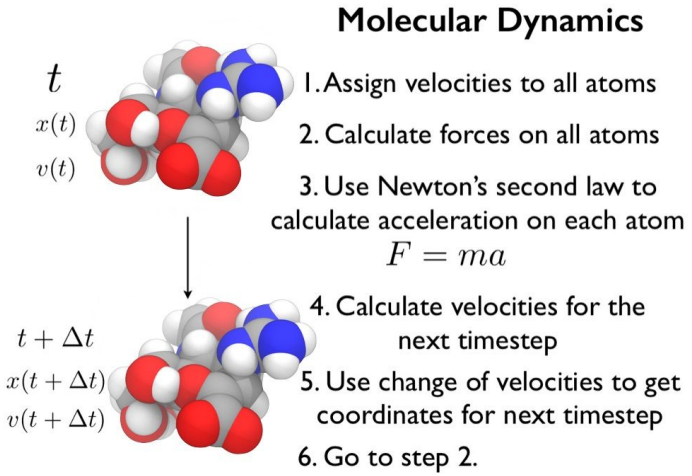
the system is now well-equilibrated at the desired temperature and pressure. We are now ready to release the position restraints and run production MD

We will run MD simulation

gmx grompp -f md.mdp -c npt.gro -t npt.cpt -p topol.top -o md_0_1.tpr

gmx mdrun -deffnm md_0_1 -s md_0_1.tpr

New files:
md_0_1.tpr: portable binary run input file. It contains the starting structure of the simulation, the molecular topology and all the simulation parameters.
md_0_1.log: *ASCII-text log file of the equilibration process*
md_0_1.cpt: *portable checkpoint file. The complete state of the simulation is stored in the checkpoint file, including extended thermostat/barostat variables, random number states and NMR time averaged data*
md_0_1.edr*: Binary energy file*
md_0_1.xtc: *Binary full-precision trajectory*
md_0_1.gro*:* starting structure of the simulation in



Analysis of calculated MD simulation

Remove PBC:

```
gmx trjconv -s md_out.tpr -f md_out.xtc -pbc nojump -o md_out_noj_noPBC.xtc <<<
"System"
```

Center system:

```
gmx trjconv -s md_out.tpr -f md_out_noj_noPBC.xtc -o md_centermolsnoPBC.xtc -pbc mol
-center -n index.ndx <<< "Protein_UNL System"
```

Alignment of all frames (Remove rotations and translations):

```
gmx trjconv -s md_out.tpr -f md_centermolsnoPBC.xtc -fit rot+trans -o md_fit.xtc -n index.ndx
<<< "Protein_UNL System"
```

Metrics

1) **Root-mean-square deviation** (**RMSD**):

the RMSD is a measure of the difference between a initial conformation of the ligand/protein conformation and the analysed conformation

Root mean square deviations in structure

gmx rms, gmx rmsdist

The *root mean square deviation* ($RMSD$) of certain atoms in a molecule with respect to a reference structure can be calculated with the program gmx rms by least-square fitting the structure to the reference structure ($t_2 = 0$) and subsequently calculating the $RMSD$ ((458)).

$$RMSD(t_1, t_2) = \left[\frac{1}{M} \sum_{i=1}^{N} m_i \|\mathbf{r}_i(t_1) - \mathbf{r}_i(t_2)\|^2 \right]^{\frac{1}{2}} \quad (458)$$

where $M = \sum_{i=1}^{N} m_i$ and $\mathbf{r}_i(t)$ is the position of atom i at time t. **Note** that fitting does not have to use the same atoms as the calculation of the $RMSD$; *e.g.* a protein is usually fitted on the backbone atoms (N, C_α, C), but the $RMSD$ can be computed of the backbone or of the whole protein.

Instead of comparing the structures to the initial structure at time $t = 0$ (so for example a crystal structure), one can also calculate (458) with a structure at time $t_2 = t_1 - \tau$. This gives some insight in the mobility as a function of τ. A matrix can also be made with the $RMSD$ as a function of t_1 and t_2, which gives a nice graphical interpretation of a trajectory. If there are transitions in a trajectory, they will clearly show up in such a matrix.

Alternatively the $RMSD$ can be computed using a fit-free method with the program gmx rmsdist:

$$RMSD(t) = \left[\frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \|\mathbf{r}_{ij}(t) - \mathbf{r}_{ij}(0)\|^2 \right]^{\frac{1}{2}} \quad (459)$$

where the *distance* $\mathbf{r}_{ij}$ between atoms at time t is compared with the distance between the same atoms at time 0.

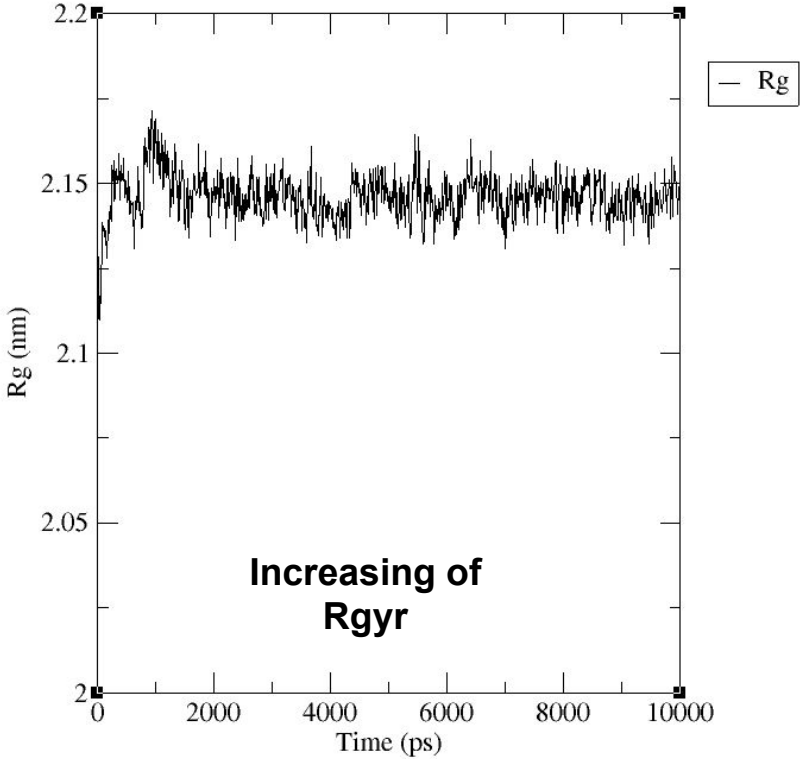
To have a rough measure for the compactness of a structure, you can calculate the *radius of gyration* with the program gmx gyrate as follows:

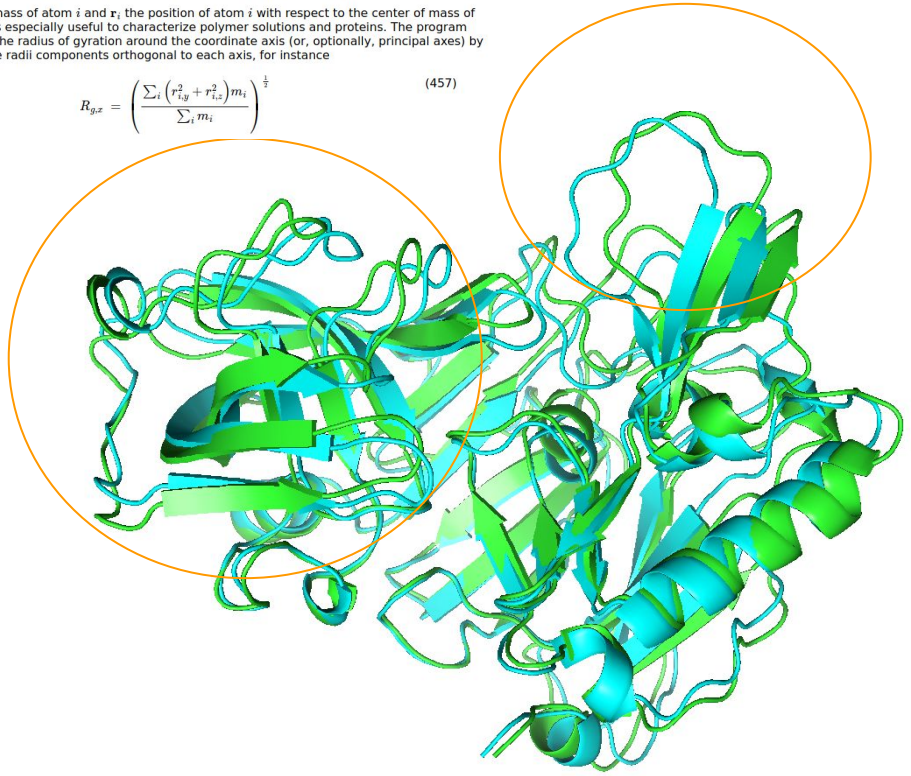
$$R_g = \left(\frac{\sum_i \|\mathbf{r}_i\|^2 m_i}{\sum_i m_i} \right)^{\frac{1}{2}} \quad (456)$$

where m_i is the mass of atom i and $\mathbf{r}_i$ the position of atom i with respect to the center of mass of the molecule. It is especially useful to characterize polymer solutions and proteins. The program will also provide the radius of gyration around the coordinate axis (or, optionally, principal axes) by only summing the radii components orthogonal to each axis, for instance

$$R_{g,z} = \left(\frac{\sum_i \left(r_{i,y}^2 + r_{i,z}^2 \right) m_i}{\sum_i m_i} \right)^{\frac{1}{2}} \quad (457)$$

Radius of gyration (total and around axes)





Conformer at the 1st ps
Conformer at the 1000th ps

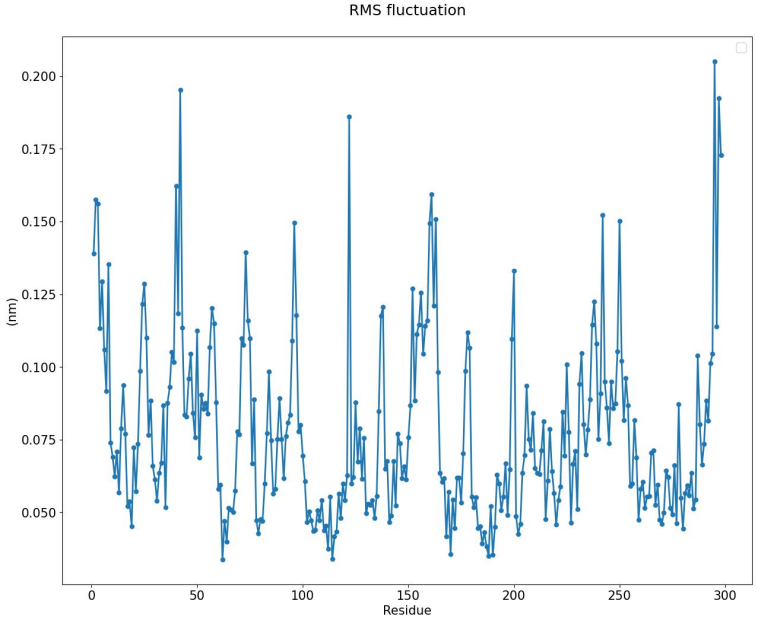
Value of Rgyr:

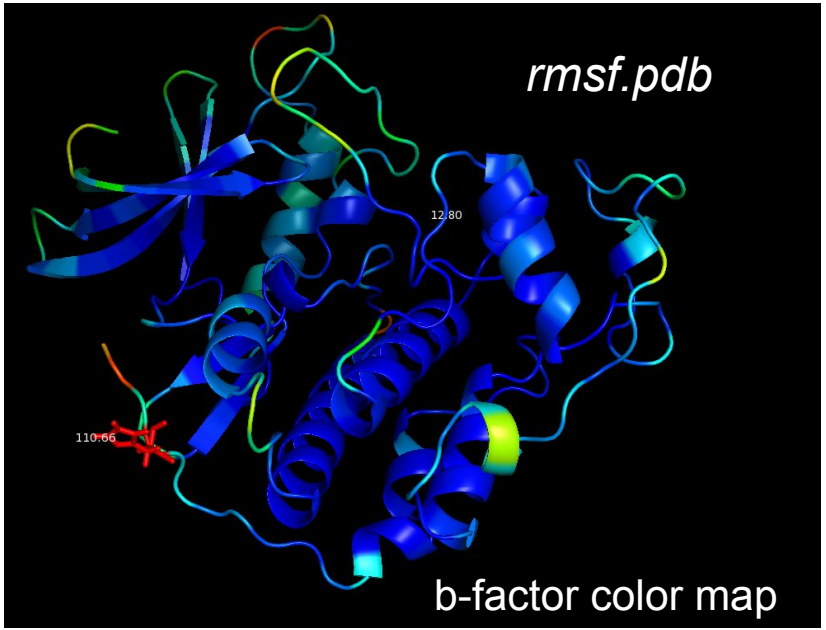
Decreasing - compression

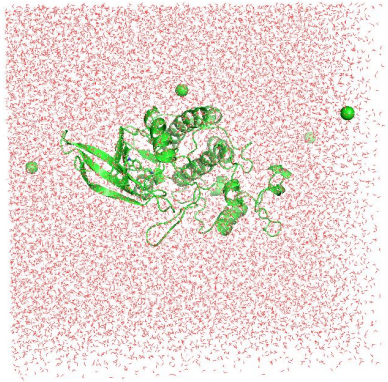
Increasing - extension

Root mean square fluctuation (RMSF, i.e. standard deviation)

`gmx rmsf` computes the root mean square fluctuation (RMSF, i.e. standard deviation) of atomic positions in the trajectory (supplied with `-f`) after (optionally) fitting to a reference frame (supplied with `-s`).

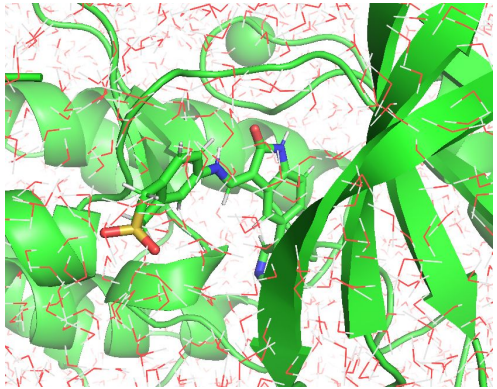






Check your own MD trajectory

frame.pdb - a frame from the trajectory to provide topology
md_fit.xtc - your fitted MD trajectory



MD simulation by one command:

Output files:

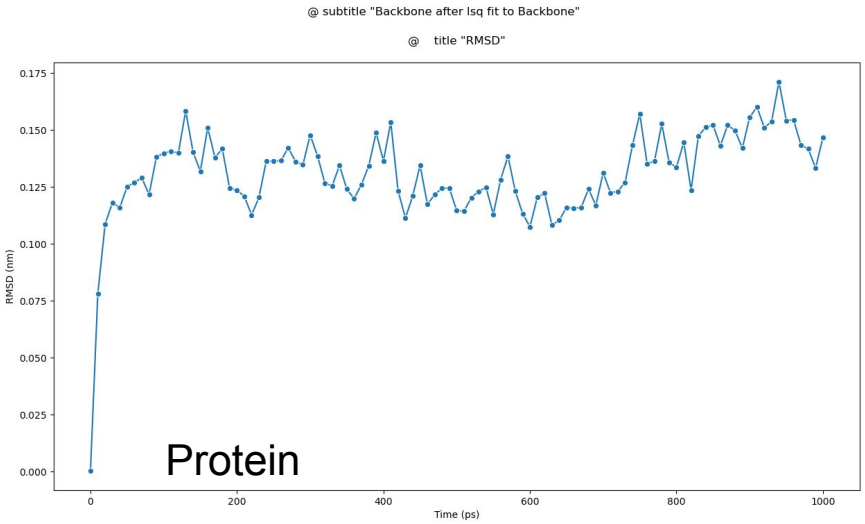
md_out.tpr: portable binary run input file. It contains the starting structure of the simulation, the molecular topology and all the simulation parameters.
md*_out.*log: *ASCII-text log file of the equilibration process*
md*_out.*cpt: *portable checkpoint file. The complete state of the simulation is stored in the checkpoint file, including extended thermostat/barostat variables, random number states and NMR time averaged data*
***md*_out.*edr*:** *Binary energy file*
md*_out.*xtc: *Binary full-precision trajectory*
***md*_out.*gro*:** starting structure of the simulation

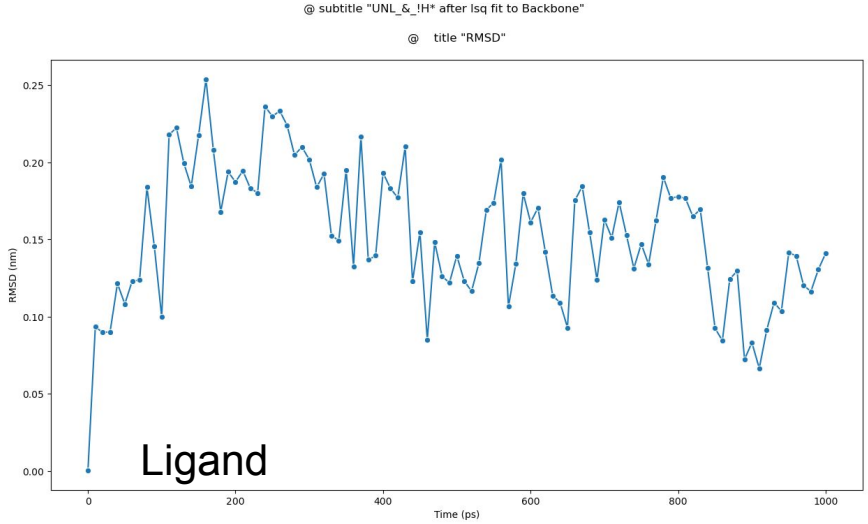
md_fit.xtc: *fitted trajectory (removed PBC and the rotation and translation, centered) to use for rmsd and energy calculation analysis*
md_short_forcheck.xtc: *fitted short trajectory (each 100 step is skipped)*
frame.pdb: a frame from the trajectory to provide topology

MD trajectory analysis files:
rmsd.png - rmsd of Backbone of protein
rmsd_ligand_1.png - rmsd of the heavy atoms of ligand
gyrate.png - radius of gyration of the protein
rmsf.png - root mean square fluctuation of each amino acids

Analysis of calculated MD simulation

```
(md) [anikonenko1@cn299.karolina MD_tutorial]$ ls mdrun/md_files/md_run/protein_HIS_ligand_1/
all.itp              em.tpr      ligand_1.itp           mdout.mdp              npt.edr  nvt.gro       pressure.png       rmsf.pdb                                                  topol.top
all_ligand_resid.txt em.trr      md_centermolsnoPBC.xtc md_out_noj_noPBC.xtc   npt.gro  nvt.log       pressure.xvg       rmsf.png
complex.gro          frame.pdb   md_fit.xtc             md_out.tpr             npt.log  nvt.mdp       rmsd_ligand_1.png  rmsf.xvg
density.png          gyrate.png  md.mdp                 md_out.xtc             npt.mdp  nvt.tpr       rmsd_ligand_1.xvg  solv.gro
density.xvg          gyrate.xvg  md_out.cpt             md_short_forcheck.xtc  npt.tpr  nvt.trr       rmsd.png           solv_ions.gro
em.edr               index.ndx   md_out.edr             minim.mdp              npt.trr  posre.itp     rmsd_xtal.png      streamd_bash_protein_HIS_ligand__26-11-2023-20-10-17.log
em.gro               ions.mdp    md_out.gro             newbox.gro             nvt.cpt  potential.png rmsd_xtal.xvg      temperature.png
em.log               ions.tpr    md_out.log             npt.cpt                nvt.edr  potential.xvg rmsd.xvg           temperature.xvg
```



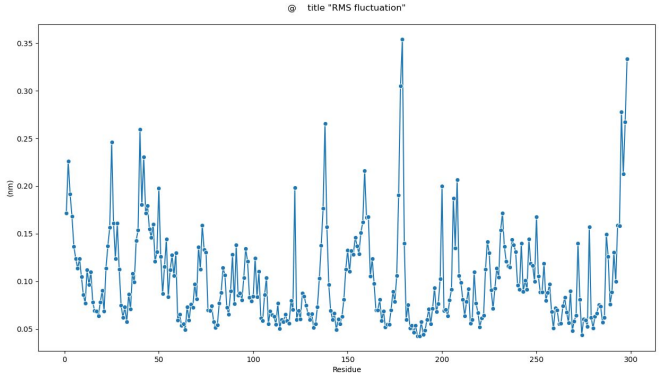


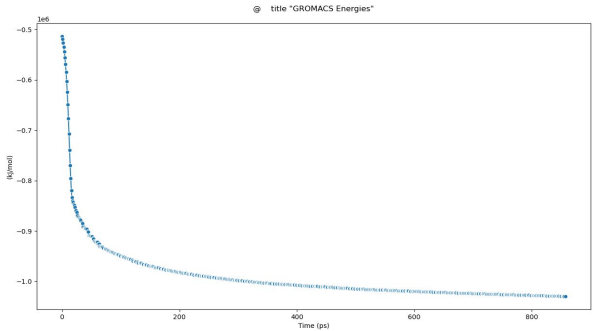
1 ns

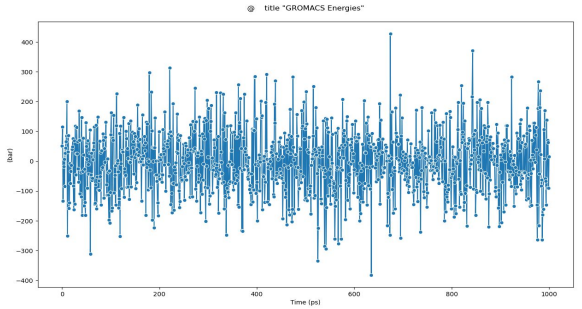
Analysis of the calculated MD simulation

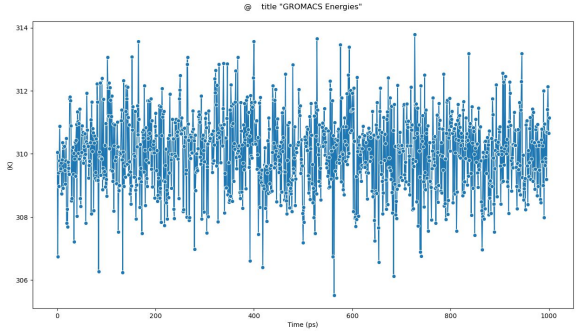
```
# This file was created Mon Nov 28 18:33:21 2022
# Created by:
#         :-) GROMACS - gmx rms, 2021.4-EasyBuild-4.5.1-PLUMED-2.7.3 (-:
#
# Executable:   /apps/all/GROMACS/2021.4-foss-2020b-PLUMED-2.7.3/bin/gmx
# Data prefix:  /apps/all/GROMACS/2021.4-foss-2020b-PLUMED-2.7.3
# Working dir:  /mnt/proj1/dd-22-84/MD_tutorial/ligand_01ns
# Command line:
#   gmx rms -s md_out.tpr -f md_fit.xtc -o rmsd.xvg -n index.ndx -tu ps
# gmx rms is part of G R O M A C S:
#
# God Rules Over Mankind, Animals, Cosmos and Such
#
@    title "RMSD"
@    xaxis  label "Time (ps)"
@    yaxis  label "RMSD (nm)"
@TYPE xy
@ subtitle "Backbone after lsq fit to Backbone"
    0.0000000    0.0004955
   10.0000000    0.0872569
   20.0000000    0.0738152
   30.0000000    0.1004452
   40.0000000    0.1048534
   50.0000000    0.0945487
   60.0000000    0.0887801
   70.0000000    0.0993429
   80.0000000    0.1069565
   90.0000000    0.1168087
  100.0000000    0.1420962
```

```
# This file was created Mon Nov 28 18:33:21 2022
# Created by:
#         :-) GROMACS - gmx rms, 2021.4-EasyBuild-4.5.1-PLUMED-2.7.3 (-:
#
# Executable:   /apps/all/GROMACS/2021.4-foss-2020b-PLUMED-2.7.3/bin/gmx
# Data prefix:  /apps/all/GROMACS/2021.4-foss-2020b-PLUMED-2.7.3
# Working dir:  /mnt/proj1/dd-22-84/MD_tutorial/ligand_01ns
# Command line:
#   gmx rms -s md_out.tpr -f md_fit.xtc -o rmsd_UNL.xvg -n index.ndx -tu ps
# gmx rms is part of G R O M A C S:
#
# God Rules Over Mankind, Animals, Cosmos and Such
#
@    title "RMSD"
@    xaxis  label "Time (ps)"
@    yaxis  label "RMSD (nm)"
@TYPE xy
@ subtitle "UNL_&_!H* after lsq fit to Backbone"
    0.0000000    0.0005219
   10.0000000    0.0578194
   20.0000000    0.0515443
   30.0000000    0.1366709
   40.0000000    0.1673483
   50.0000000    0.1613055
   60.0000000    0.1436562
   70.0000000    0.1480425
   80.0000000    0.1522363
   90.0000000    0.1496438
  100.0000000    0.0998629
```







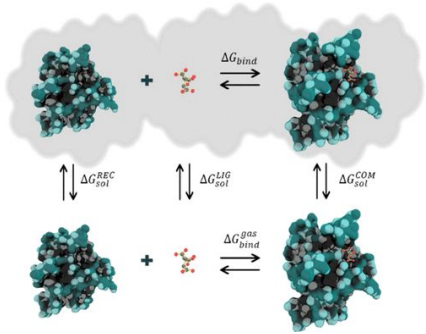


MMPBSA / MMGBSA

End-state free energy calculations
with GROMACS files

What can be done by MD

- to estimate binding affinity of protein-ligand complexes
  - Molecular mechanics Poisson–Boltzmann surface area (**MM/PBSA**)
  - Molecular mechanics generalized Born surface area (**MM/GBSA**)



In the MM/PBSA or MM/GBSA approach, the free energy for binding of the ligand (L) to the protein receptor (R) to form the complex (RL),

$$\Delta G_{\text{bind}} = G_{\text{RL}} - G_{\text{R}} - G_{\text{L}} \quad (4)$$

can be decomposed into contributions of different interactions and expressed as (58)

$$\Delta G_{\text{bind}} = \Delta H - T\Delta S = \Delta E_{\text{MM}} + \Delta G_{\text{sol}} - T\Delta S \quad (5)$$

in which

$$\Delta E_{\text{MM}} = \Delta E_{\text{int}} + \Delta E_{\text{ele}} + \Delta E_{\text{vdW}} \quad (6)$$

$$\Delta G_{\text{sol}} = \Delta G_{\text{PB/GB}} + \Delta G_{\text{SA}} \quad (7)$$

$$\Delta G_{\text{SA}} = \gamma \cdot \text{SASA} + b \quad (8)$$

Valdés-Tresanco, M.S., Valdés-Tresanco, M.E., Valiente, P.A. and Moreno E. ***gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS.*** Journal of Chemical Theory and Computation, 2021 17 (10), 6281-6291. https://pubs.acs.org/doi/10.1021/acs.jctc.1c00645.
***MMPBSA.py*: An Efficient Program for End-State Free Energy Calculations** Bill R. Miller III, T. Dwight McGee Jr., Jason M. Swails, Nadine Homeyer, Holger Gohlke, and Adrian E. Roitberg Journal of Chemical Theory and Computation **2012** *8* (9), 3314-3321 DOI: 10.1021/ct300418h

What can be done by MD

- to estimate binding affinity of protein-ligand complexes

Total $G_{Binding}$=

- **Gas-phase molecular mechanics energy ΔE_{MM}:**
  - changes in the **internal energies** ΔE_{int} (bond, angle, and dihedral energies)
  - **electrostatic energies** ΔE_{ele}
  - **van der Waals energies** ΔE_{vdW}
- **sum of the electrostatic solvation energy G_{sol}**
  - **The polar contribution** is calculated using either the PB or GB model (ΔGPB/GB). Where the GB method gives an analytical expression for the polar solvation energy and is thus much faster than the PB method.
  - **nonpolar energy** is usually estimated using the solvent-accessible surface area (**SASA**)
- The change in **conformational entropy** **$-T\Delta S$**
  - is usually calculated by normal-mode analysis (or Interaction entropy) on a set of conformational snapshots taken from MD simulations.

In the MM/PBSA or MM/GBSA approach, the free energy for binding of the ligand (L) to the protein receptor (R) to form the complex (RL),

$$\Delta G_{bind} = G_{RL} - G_R - G_L \quad (4)$$

can be decomposed into contributions of different interactions and expressed as (58)

$$\Delta G_{bind} = \Delta H - T\Delta S = \Delta E_{MM} + \Delta G_{sol} - T\Delta S \quad (5)$$

in which

$$\Delta E_{MM} = \Delta E_{int} + \Delta E_{ele} + \Delta E_{vdW} \quad (6)$$

$$\Delta G_{sol} = \Delta G_{PB/GB} + \Delta G_{SA} \quad (7)$$

$$\Delta G_{SA} = \gamma \cdot SASA + b \quad (8)$$

MMPBSA.in

```
Sample input file for PB/GB calculation
#This input file is meant to show only that gmx_MMPBSA works. Although, we tried to use the input
files as recommended in the
#Amber manual, some parameters have been changed to perform more expensive calculations in a
reasonable amount of time. Feel free to change the parameters
#according to what is better for your system.
&general
sys_name="PB_GB_IE",
startframe=1, interval=1, verbose=2, PBRadii=3,
interaction_entropy=1, ie_segment=100, temperature=310
/
&gb
igb=5, saltcon=0.150,
/
&pb
istrng=0.15, fillratio=4.0, radiopt=0, indi=1, exdi=80.0
/
```

run_gbsa -i mdrun/md_files/md_run/protein_HIS_ligand_1/

```
---------------------------------------------------------------------------------
Ligand:
Energy Component       Average     SD(Prop.)         SD   SEM(Prop.)        SEM
---------------------------------------------------------------------------------
BOND                     10.44          1.76       1.76         0.53       0.53
ANGLE                    45.94          3.47       3.47         1.05       1.05
DIHED                    22.15          2.81       2.81         0.85       0.85
VDWAALS                  -3.17          0.66       0.66         0.20       0.20
EEL                      95.19          1.06       1.06         0.32       0.32
1-4 VDW                   8.18          0.73       0.73         0.22       0.22
1-4 EEL                -225.44          1.28       1.28         0.38       0.38
EGB                     -36.44          1.07       1.07         0.32       0.32
ESURF                     3.63          0.02       0.02         0.01       0.01

GGAS                    -46.70          5.17       4.07         1.56       1.23
GSOLV                   -32.81          1.07       1.07         0.32       0.32

TOTAL                   -79.51          5.28       4.12         1.59       1.24

---------------------------------------------------------------------------------
---------------------------------------------------------------------------------
Delta (Complex - Receptor - Ligand):
Energy Component       Average     SD(Prop.)         SD   SEM(Prop.)        SEM
---------------------------------------------------------------------------------
ΔBOND                    -0.00          0.83       0.00         0.25       0.00
ΔANGLE                   -0.00          2.81       0.00         0.85       0.00
ΔDIHED                    0.00          2.53       0.00         0.76       0.00
ΔVDWAALS                -45.23          0.57       2.79         0.17       0.84
ΔEEL                    -37.12          0.26       6.03         0.08       1.82
Δ1-4 VDW                  0.00          0.54       0.00         0.16       0.00
Δ1-4 EEL                  0.00          0.45       0.00         0.14       0.00
ΔEGB                     49.63          0.19       3.93         0.06       1.18
ΔESURF                   -6.08          0.01       0.11         0.00       0.03

ΔGGAS                   -82.35          0.62       5.83         0.19       1.76
ΔGSOLV                   43.55          0.19       3.94         0.06       1.19

ΔTOTAL                  -38.80          0.65       3.12         0.20       0.94
---------------------------------------------------------------------------------
---------------------------------------------------------------------------------
Using Interaction Entropy Approximation:
ΔG binding =     -35.81 +/-     3.31
---------------------------------------------------------------------------------
```

PBSA

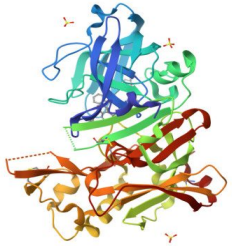
	A	B	C
	fname	**ΔG_binding**	**ΔG_binding_+/-**
	protein_HIS_igand_1	-35.81	0.74

GBSA

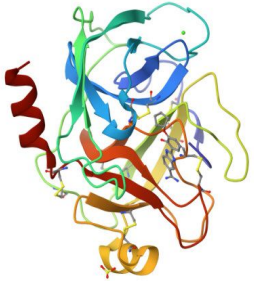
fname	**ΔG_binding**	**ΔG_binding_+/-**
protein_HIS_ligand_1	-24.31	4.1

MMPBSA Energy and **MMGBSA Energy** cannot be compared within the different methods. But you can rank your molecules by energies obtained from each method separately.

INSTITUTE OF MOLECULAR AND TRANSLATIONAL MEDICINE



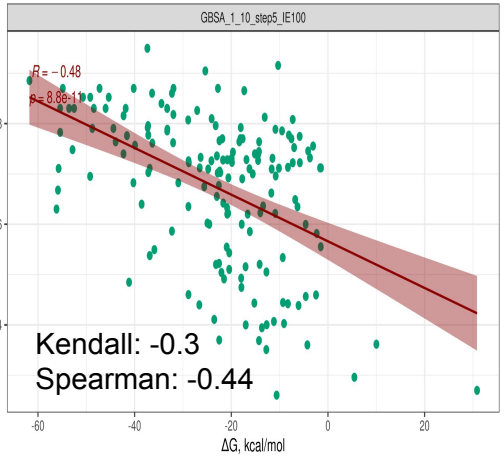
Human Beta-secretase 1
(P56817) 3UFL
165 molecules
from PDB complexes

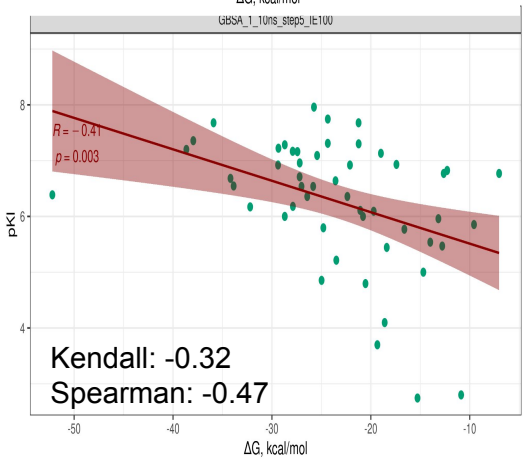


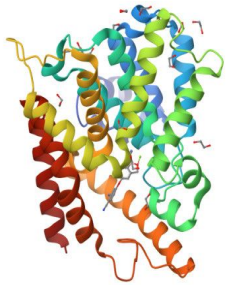
Bos taurus Beta-trypsin
(P00760) 1O2I
51 molecules
from PDB complexes

run_md module

10 ns simulation. 310 K



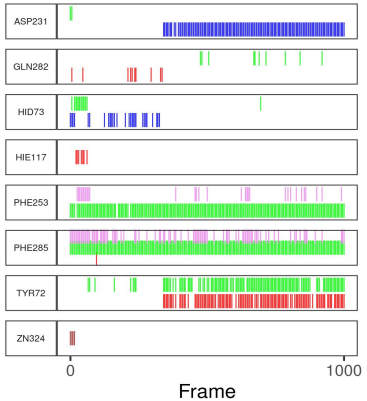




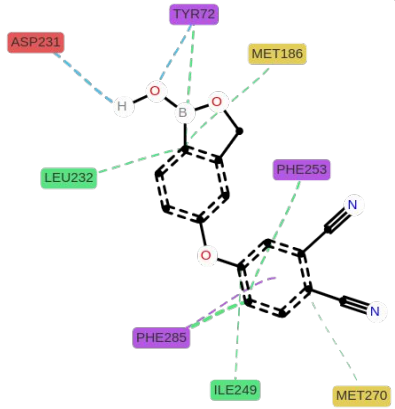
Boron-containing molecules

Human
cAMP-specific 3',5'-cyclic
phosphodiesterase 4B
PDE4B
(Q07343)
3O0J
Ki: 65 (nM)
GBSA: -23.9 kcal/mol

run_prolif module



show residue if other than just a hydrophobic interaction occurs



Acidic | Aliphatic | Aromatic | Sulfur
FaceToFace | HBDonor/HBAcceptor | Hydrophobic

Thank you for your attention!