

# 6th Advanced in silico Drug Design workshop/challenge 2023

## High-throughput MD tutorial

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



# PuTTY

**PuTTY Configuration**

Category:

- Session
  - Logging
- Terminal
  - Keyboard
  - Bell
  - Features
- Window
  - Appearance
  - Behaviour
  - Translation
- Selection
  - Colours
- Connection
  - Data
  - Proxy
  - SSH
    - Serial
    - Telnet
    - Rlogin
    - SUPDUP

**Basic options for your PuTTY session**

Specify the destination you want to connect to

Host Name (or IP address)  Port

Connection type:  
☒ SSH ☐ Serial ☐ Other:

Load, save or delete a stored session

Saved Sessions

Default Settings
------------------

Load Save Delete

Close window on exit  
☐ Always ☐ Never ☒ Only on clean exit

About Open Cancel

**PuTTY Configuration**

Category:

- Logging
- Terminal
  - Keyboard
  - Bell
  - Features
- Window
  - Appearance
  - Behaviour
  - Translation
- Selection
  - Colours
- Connection
  - Data
  - Proxy
  - SSH
    - Key
    - Host keys
    - Cipher
    - Auth
      - Credentials
      - GSSAPI
    - TTY
    - X11
    - Tunnels

**Credentials to authenticate with**

Public-key authentication

Private key file for authentication:  
 Browse...

Certificate to use with the private key:  
 Browse...

Plugin to provide authentication responses

Plugin command to run

SSH->Auth->Credentials

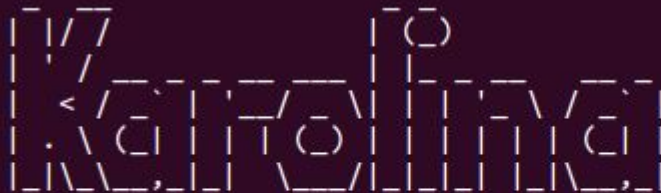
About Open Cancel



```
~$ cp ./dd-23-13-01 ~/.ssh/
```

```
~$ chmod 600 ~/.ssh/dd-23-13-01
```

```
~$ ssh -i ~/.ssh/dd-23-13-01 dd-23-13-01@login1.karolina.it4i.cz
```

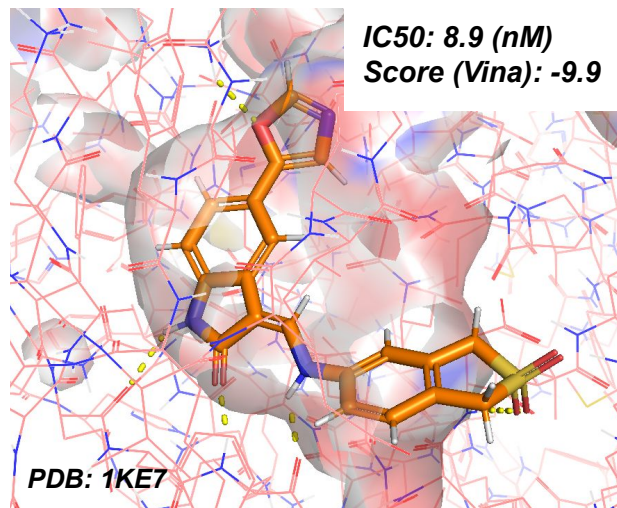
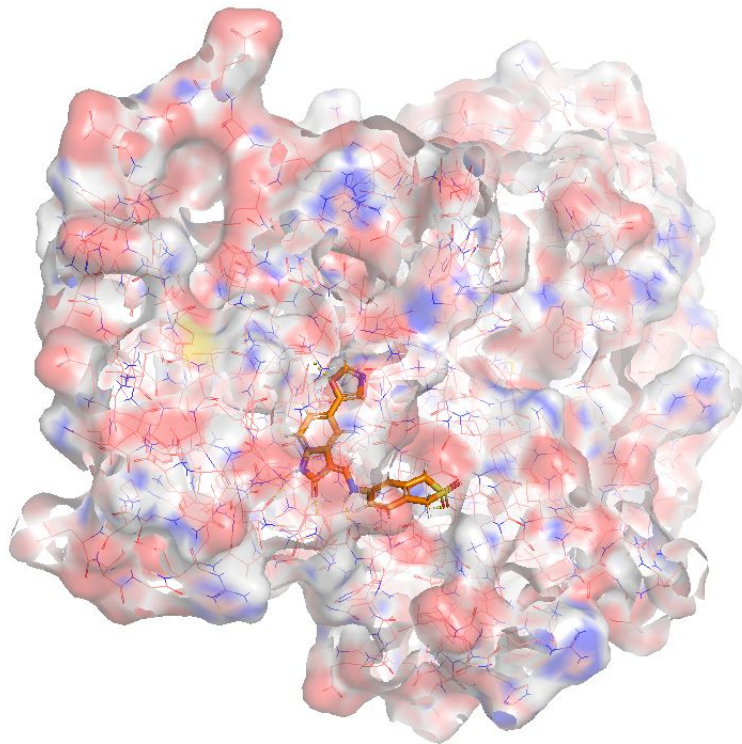


...running on Red Hat Enterprise Linux 7.x

```
[dd-23-13-01@login1.karolina ~]$
```



# Molecular Docking



**Score** – the strength of binding of the ligand and the receptor.

**Pose** – a possible relative orientation of a ligand and a receptor as well as conformation of a ligand and a receptor when they are form complex





# Automated scripts for Molecular Docking

<https://github.com/ci-lab-cz/docking-scripts>

## Installation

```
pip install moldock
```

or the latest version

```
pip install git+https://github.com/ci-lab-cz/docking-scripts.git
```

## Dependencies

from conda

```
conda install -c conda-forge python=3.9 numpy=1.20 rdkit scipy dask distributed vina
```

from pypi

```
pip install meeko
```

Installation of gnina is described at <https://github.com/gnina/gnina>



# Automated scripts for Molecular Docking

<https://github.com/ci-lab-cz/docking-scripts>

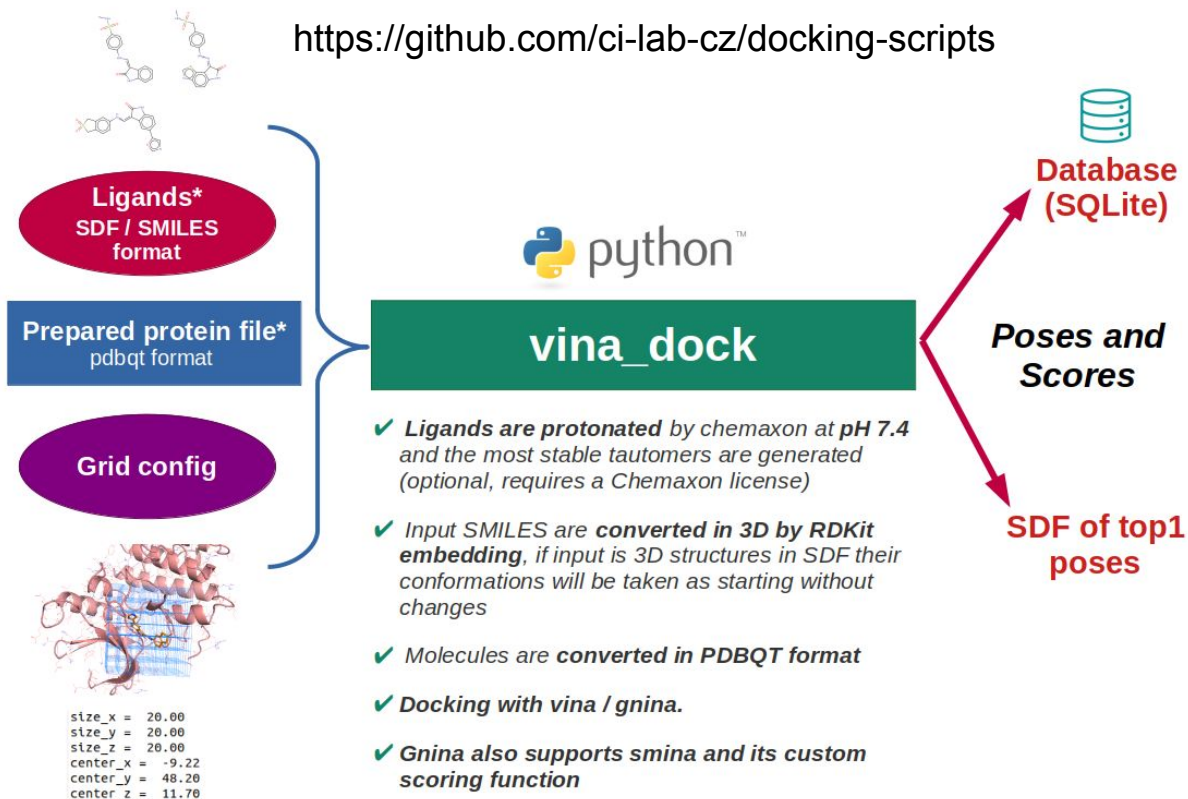




Table: mol

	id	smi	smi_protonated	source_mol_block	source_mol_block_protonated	docking_score	pdb_block	mol_block	time
	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter
1	1BJU_ligand	NC(=[NH2+])c1cc...	NC(=[NH2+])c1ccc(NC(=O)...	1BJU_ligand...	1BJU_ligand...	-6.863	MODEL 1...	1BJU_ligand...	2023-01-23 ...
2	1BJV_ligand	NC(=[NH2+])c1cc...	NC(=[NH2+])c1ccc(NC(=O)...	1BJV_ligand...	1BJV_ligand...	-7.524	MODEL 1...	1BJV_ligand...	2023-01-23 ...
3	1C5Q_ligand	NC(=[NH2+])c1cc...	NC(=[NH2+])c1cc2c(l)cccc2...	1C5Q_ligand...	1C5Q_ligand...	-6.111	MODEL 1...	1C5Q_ligand...	2023-01-23 ...
4	1C5S_ligand	NC(=[NH2+])c1cc...	NC(=[NH2+])c1cc2ccccc2s1	1C5S_ligand...	1C5S_ligand...	-5.889	MODEL 1...	1C5S_ligand...	2023-01-23 ...
5	1C5T_ligand	NC(=[NH2+])c1cc...	NC(=[NH2+])c1cc2ccccc2s1	1C5T_ligand...	1C5T_ligand...	-6.156	MODEL 1...	1C5T_ligand...	2023-01-23 ...
6	1F0T_ligand	NC(=[NH2+])c1cc...	NC(=[NH2+])c1ccc(O)c(CN2...	1F0T_ligand...	1F0T_ligand...	-7.81	MODEL 1...	1F0T_ligand...	2023-01-23 ...
7	1F0U_ligand	COC(=O)C(Cc1cc...	COC(=O)[C@H]...	1F0U_ligand...	1F0U_ligand...	-7.134	MODEL 1...	1F0U_ligand...	2023-01-23 ...
8	1G36_ligand	Cc1nc2ccccc2n1C...	Cc1nc2ccccc2n1Cc1ccc2c(c...	1G36_ligand...	1G36_ligand...	-8.78	MODEL 1...	1G36_ligand...	2023-01-23 ...
9	1GHZ_ligand	NC(=[NH2+])c1cc...	NC(=[NH2+])c1ccc2[nH]c(-...	1GHZ_ligand...	1GHZ_ligand...	-8.197	MODEL 1...	1GHZ_ligand...	2023-01-23 ...
10	1GI1_ligand	NC(=[NH2+])c1cc...	NC(=[NH2+])c1ccc2[nH]c(-...	1GI1_ligand...	1GI1_ligand...	-7.932	MODEL 1...	1GI1_ligand...	2023-01-23 ...
11	1GI6_ligand	NC(=[NH2+])c1cc...	NC(=[NH2+])c1ccc2[nH]c(-...	1GI6_ligand...	1GI6_ligand...	-8.263	MODEL 1...	1GI6_ligand...	2023-01-23 ...
12	1GJ6_ligand	NC(=[NH2+])c1cc...	NC(=[NH2+])c1cc2cc(-...	1GJ6_ligand...	1GJ6_ligand...	-8.468	MODEL 1...	1GJ6_ligand...	2023-01-23 ...
13	1K1I_ligand	NC(=[NH2+])c1cc...	NC(=[NH2+])c1cccc(C[C@H...	1K1I_ligand...	1K1I_ligand...	-8.946	MODEL 1...	1K1I_ligand...	2023-01-23 ...
14	1K1J_ligand	COC(=O)C1CCN(C...	COC(=O)C1CCN(C(=O)[C@...	1K1J_ligand...	1K1J_ligand...	-7.579	MODEL 1...	1K1J_ligand...	2023-01-23 ...
15	1K1L_ligand	NC(=[NH2+])c1cc...	NC(=[NH2+])c1cccc(C[C@H...	1K1L_ligand...	1K1L_ligand...	-6.695	MODEL 1...	1K1L_ligand...	2023-01-23 ...

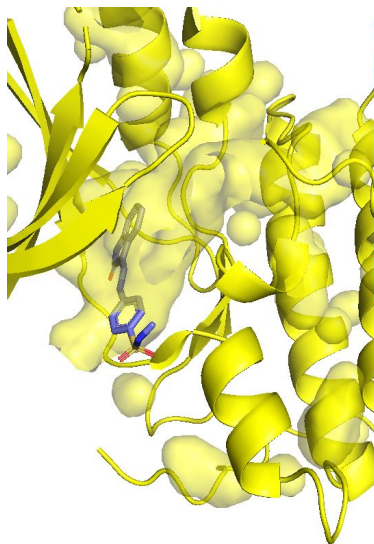
Table: setup

protonation	protonation_done	protein_pdbqt	protein_setup
Filter	Filter	Filter	Filter
1	1	ATOM 1 ...	size_x = ...

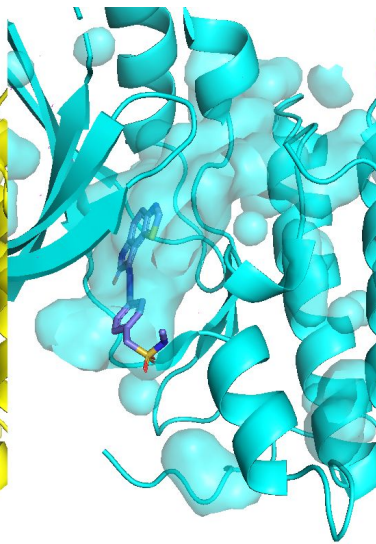
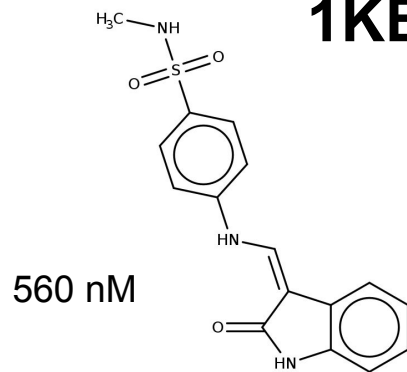
Output SQLite DB



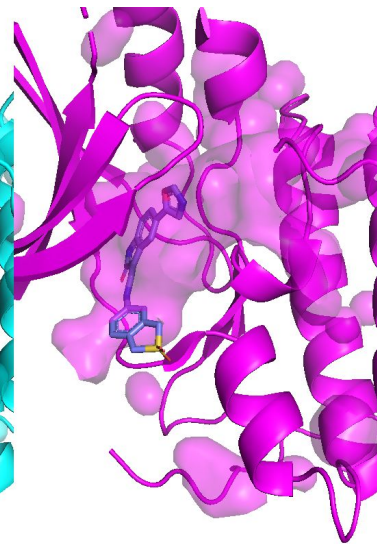
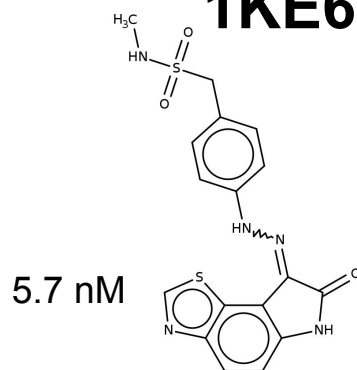
**CYCLIN-DEPENDENT  
KINASE 2 (CDK2)**



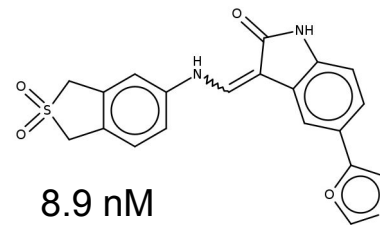
**1KE5**



**1KE6**



**1KE7**





# Automated scripts for Molecular Docking

```
ssh -i ~/.ssh/dd-23-13-01 dd-23-13-01@login1.karolina.it4i.cz
```

```
mkdir docking_tutorial
```

```
cd docking_tutorial
```

```
cp
```

```
/mnt/proj2/dd-23-13/docking_tutorial/{ligands_pH74.smi,protein_prepared.pdbqt,config.txt} .
```

```
qsub -A DD-23-13 -v
```

```
input=$(pwd)/ligands.smi,protein=$(pwd)/protein_prepared.pdbqt,config=$(pwd)/config.txt
```

```
/mnt/proj2/dd-23-13/script_vina.pbs
```





# Prepare docked molecules for MD

```
module load Anaconda3
```

```
source activate gmxMMPBSA
```

```
python /mnt/proj2/dd-23-13/md-scripts/scripts/sdf2mols.py -i
```

```
docking_tutorial/ligands_pH74_protein_prepared_docking_vina_output.sdf -o
```

```
docking_tutorial/mols
```



# High-throughput MD



## Run MD simulation by one command:

```
ssh -i ~/.ssh/dd-23-13-01 dd-23-13-01@login1.karolina.it4i.cz
```

```
mkdir md_tutorial
```

```
cd md_tutorial
```

```
cp /mnt/proj2/dd-23-13/md_tutorial/{protein_prepared.pdb,ligand.mol} .
```

```
qsub -A DD-23-13 -v
```

```
lfile=$(pwd)/ligand.mol,pfile=$(pwd)/protein_prepared.pdb,script_path=/mnt/proj2/dd-23-13/md-scripts/scripts  
/,wdir=ligand_01ns,mdtime=0.1,gromacs_version='GROMACS/2021.4-foss-2020b-PLUMED-2.7.3'  
/mnt/proj2/dd-23-13/md-scripts/Protein-Ligand_MD_PBGBSA/01_complex_preparation_md.pbs
```

### in case of multiple ligands:

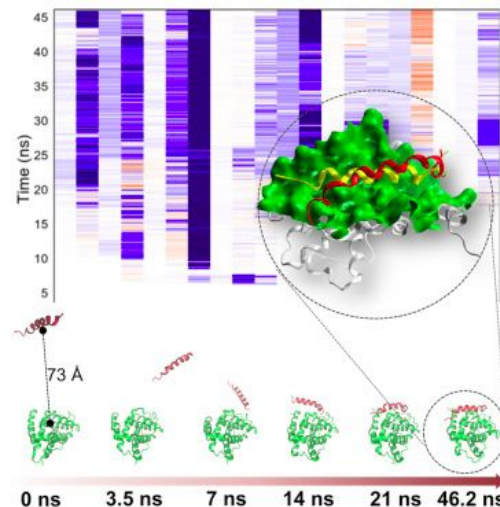
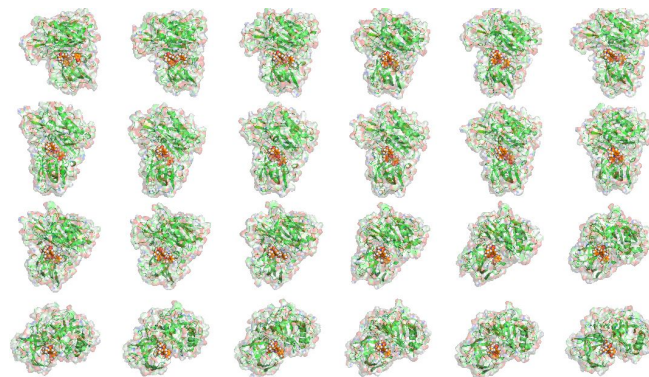
```
for i in mols/*.mol;do fname=${i##*/}; name=${fname%.*}; echo $i; qsub -A DD-23-13 -v  
lfile=$(pwd)/$i,pfile=$(pwd)/protein_prepared.pdb,script_path=/mnt/proj2/dd-23-13/md-scripts/scripts/,wdir=$(  
pwd)/$name,mdtime=0.1,gromacs_version='GROMACS/2021.4-foss-2020b-PLUMED-2.7.3'  
/mnt/proj2/dd-23-13/md-scripts/Protein-Ligand_MD_PBGBSA/01_complex_preparation_md.pbs;done
```



# 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.**

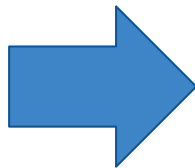
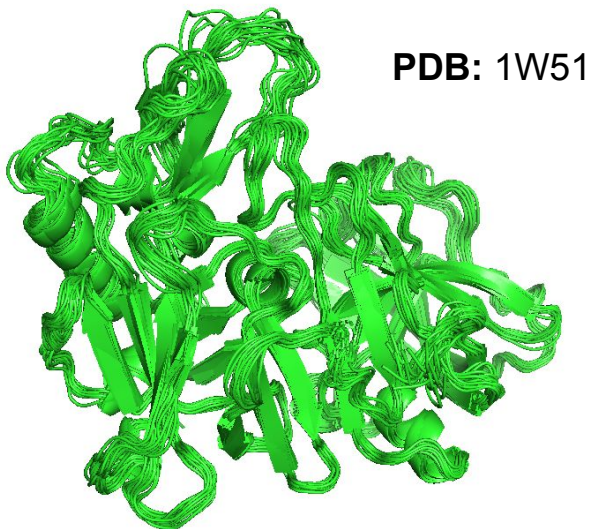
- to explore conformational space
- to explore biological process of molecular recognition
- to investigate ligand pose stability
- to estimate binding affinity of protein-ligand complexes



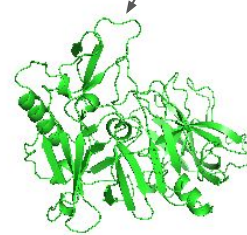
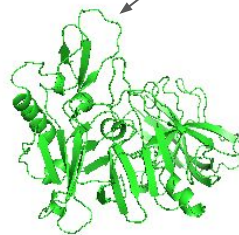
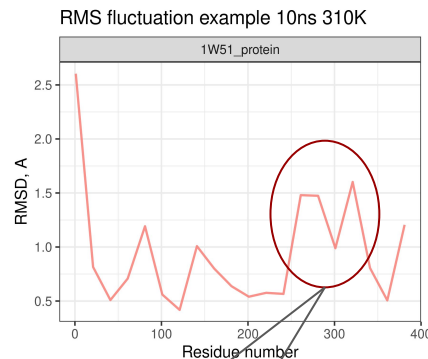
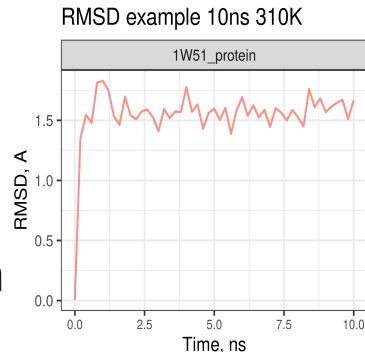


# What can be done by MD

- 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



Conformation can be used:  
- for further molecular docking study



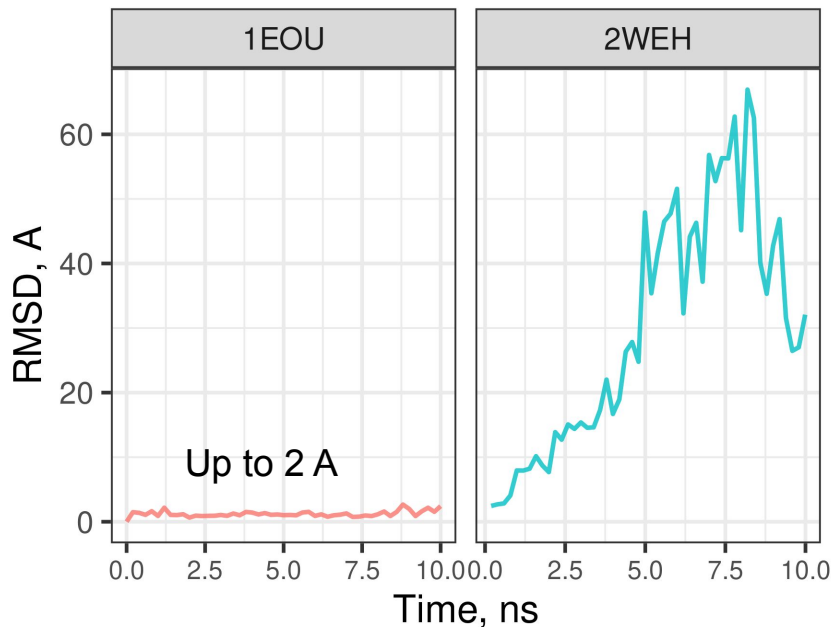




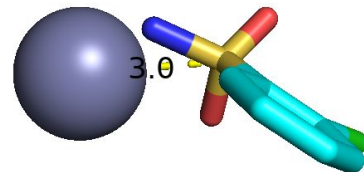
# What can be done by MD

- To explore stability of ligand pose

RMSD example 10ns 310K

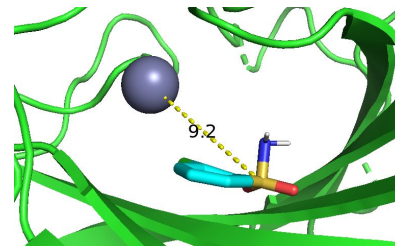


## Example of incorrect protonation and pose:

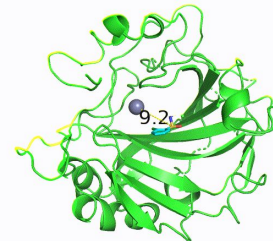


x-ray

PDB: 2WEH



docking

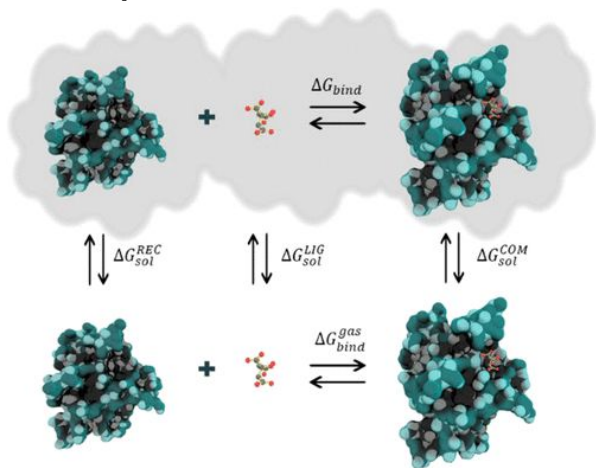


**MD 10ns.**  
*Ligand goes out of site*

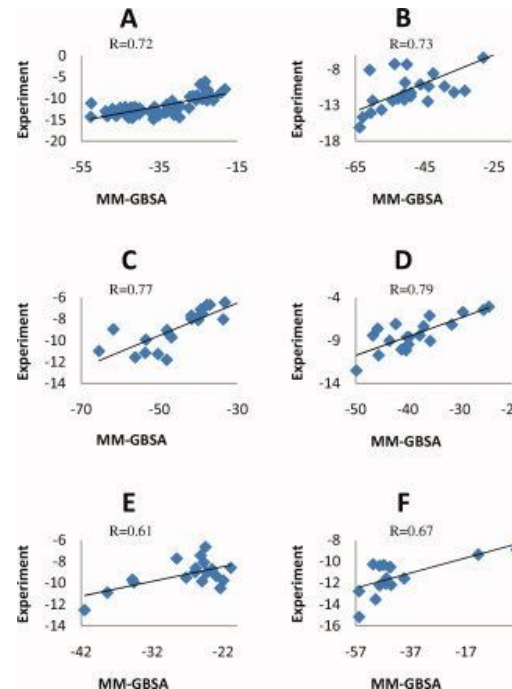


# What can be done by MD

- 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

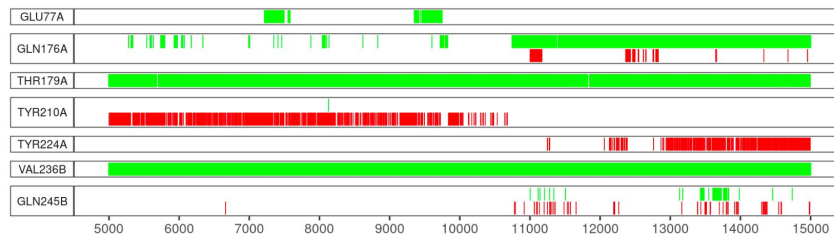
- to investigate protein-ligand interaction stability

[5]:

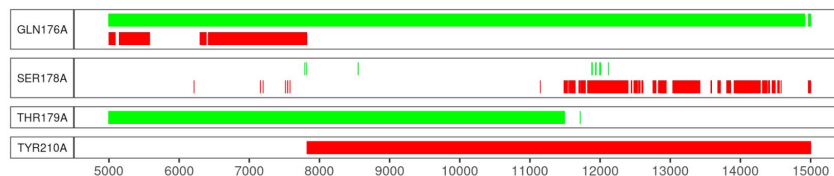
ligand	LIG1.G						
protein	TYR38.A		TYR109.A		THR110.A		TRP125.A
interaction	Hydrophobic	VdWContact	Hydrophobic	VdWContact	Hydrophobic	Hydrophobic	VdWContact
Frame							
0	False	False	True	False	False	True	False
10	False	False	True	True	False	True	False
20	False	False	True	True	False	True	True
30	True	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>

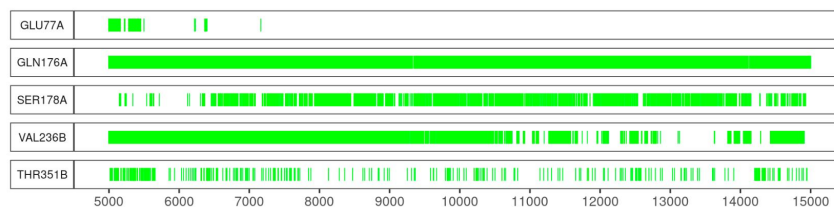
A.



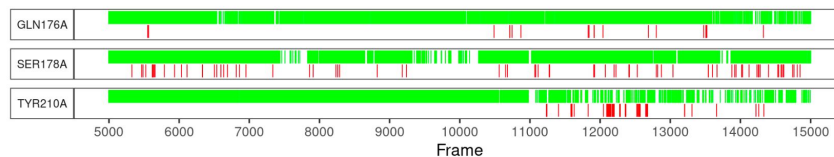
B.



C.

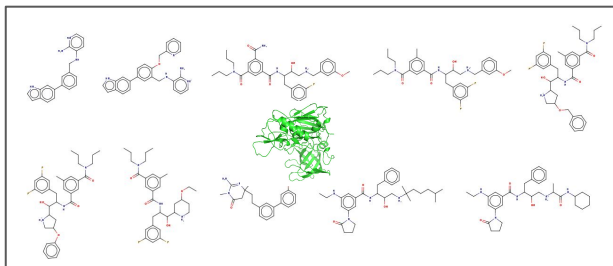


D.





**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

**Preprocessing**

**Pre-simulation  
steps**

## Structure Preparation

- add missing atoms
- protonation
- etc

## Force-field

- Definitions of inter-atomic bonded and non-bonded forces

## Simulation box setup

- box size/shape

## Solvate system

- add HOH molecules

## Neutralize system

- add NA+/CL- ions

## Energy minimization

- 100 ps

## NVT and NPT equilibration

- 1000 ps

## Product Simulation

## Analysis



ci-lab-cz / md-scripts Public

<> Code Issues 4 Pull requests Actions Projects Security Insights

master md-scripts / Protein-Ligand\_MD\_PGBSA /

Go to file Add file ...

avnikonenko and DrrDom Update README.md 35e6bef 13 hours ago History

01_complex_preparation_md.pbs	Update 01_complex_preparation_md.pbs	2 months ago
02_pbsa.pbs	Remove redundant code	3 months ago
03_sum_result_pbsa.pbs	New version of gmx_PBSA changes	3 months ago
README.md	Update README.md	13 hours ago

README.md

## Protein-ligand molecular dynamics simulation + PB(GB)SA calculation

Scripts:

- 01\_complex\_preparation\_md.pbs
- 02\_pbsa.pbs
- 03\_sum\_result\_pbsa.pbs

***The tool is already implemented  
and free available***

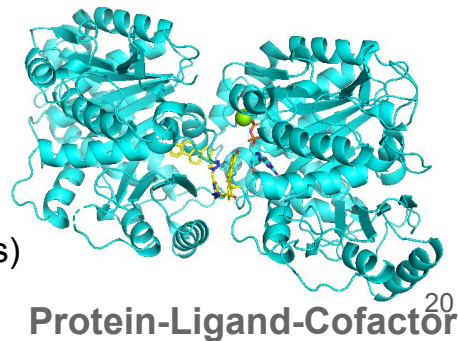
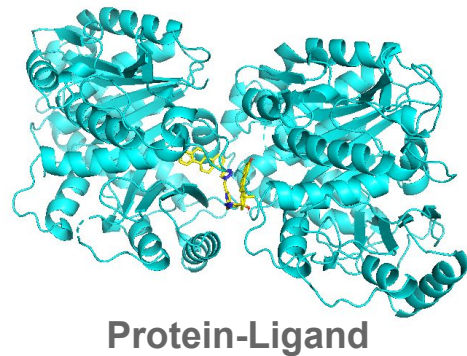
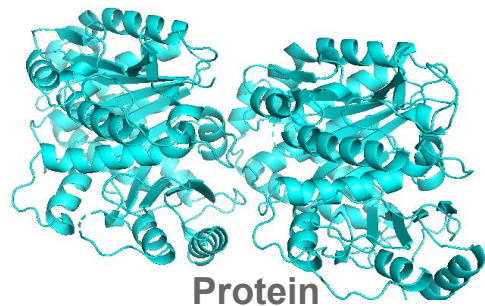
Command: `qsub -v lfile=ligand.mol,pfile=protein.pdb,script_path=/scripts,wdir=ligand_protein_1ns,mdtime=1 01_complex_preparation_md.pbs`





# Main features of the tool:

- **User control of simulation time**
  - from 10 ps to 1  $\mu$ s
- **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
- **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





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

RCSB PDB PROTEIN DATA BANK 196,528 Structures from the PDB 1,000,361 Computed Structure Models (CSM)

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

Advanced Search Browse Annotations

PDB-101 PDB PDDataResource Nucleic Acid Database wwPDB Foundation

Structure Summary 3D View Annotations Experiment Sequence Genome Ligands Versions

Biological Assembly 1



1KE7

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

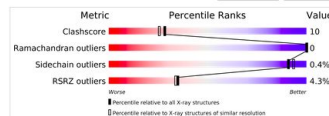
Method: X-RAY DIFFRACTION

Resolution: 2.00 Å

R-Value Free: 0.235

R-Value Work: 0.191

wwPDB Validation



Find Similar Assemblies

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)	

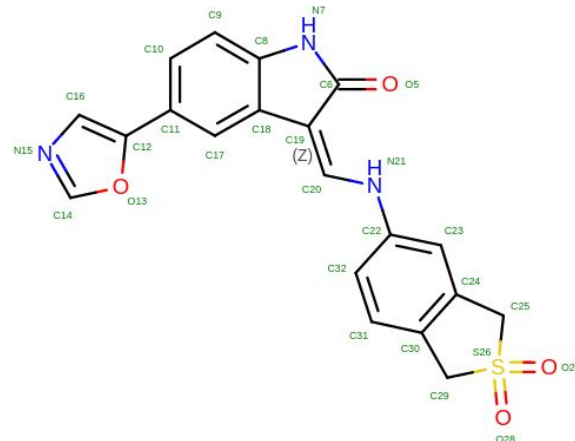
UniProt & NIH Common Fund Data Resources

Find proteins for P24941 (Homo sapiens)

Explore P24941

Go to UniProtKB: P24941

1KE7: Ligand LS3



Small Molecules

Ligands (1 Unique)

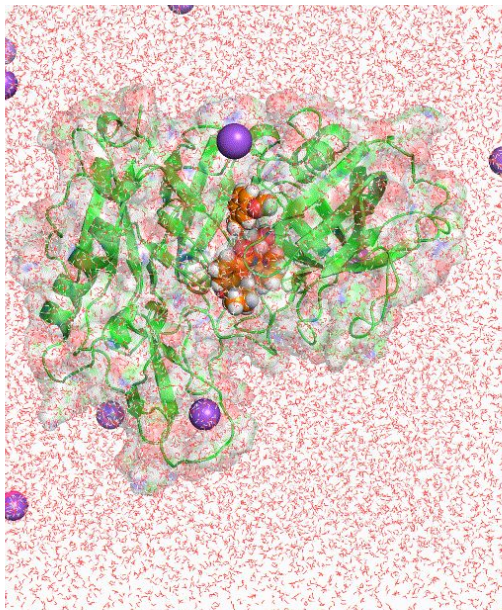
ID	Chains	Name / Formula / InChI Key	2D Diagram	3D Interactions
LS3	B [auth A]	3-[[[(2,2-DIOXIDO-1,3-DIHYDRO-2-BENZOTHIEN-5-YL)AMINO]METHYLENE]-5-(1,3-OXAZOL-5-YL)-1,3-DIHYDRO-2H-INDOL-2-ONE C <sub>29</sub> H <sub>15</sub> N <sub>3</sub> O <sub>4</sub> S FTQVGMRLRLXBPT-IDUWGFVSA-N		Ligand Interaction

Binding Affinity Annotations

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



# Classical Molecular Dynamics



Preprocessing

Pre-simulation  
steps

## Structure Preparation

### 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

### Energy minimization

- 100 ps
- to ensure that the system has no steric clashes or inappropriate geometry

### NVT and NPT equilibration

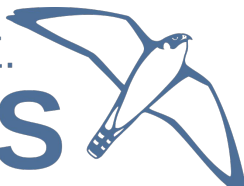
- 1000 ps
- equilibrate the solvent and ions around the protein

## Product Simulation

## Analysis


FAST. FLEXIBLE. FREE.

# GROMACS





# Protein preparation



about projects people publications  
resources visit us search

## UCSF CHIMERA

an Extensible Molecular Modeling System

UCSF Chimera is a program for the interactive visualization and analysis of molecular structures and related data, including density maps, trajectories, and sequence alignments. It is available free of charge for noncommercial use. Commercial users, please see [Chimera commercial licensing](#).

**We encourage Chimera users to try ChimeraX** for much better performance with large structures, as well as other major [advantages](#) and completely new features. ChimeraX includes a significant subset of Chimera features (with more to come, see the [missing features list](#)) and is under active development. Users may choose to use both programs, and it is fine to have both installed.

**Chimera is no longer under active development**, and is only updated for critical maintenance. Chimera development was supported by a grant from the [National Institutes of Health](#) (P41-GM103311) that ended in 2018.

### Quick Links

[Documentation](#)  
[Getting Started](#)  
[User's Guide](#)  
[Command Index](#)  
[Tutorials and Videos](#)  
[Guide to Volume Data](#)  
[Release Notes](#)  
[Download](#)  
[What's New in Daily Builds](#)  
[Map of Download Locations](#)  
[Galleries](#)  
[Image Gallery](#)  
[Animation Gallery](#)  
[Publications and Talks](#)  
[Related Databases and Software](#)  
[Citing Chimera](#)  
[Contact Us](#)

### Recent Citations

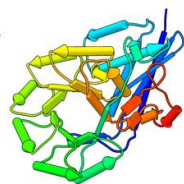
[Imprinted antibody responses against SARS-CoV-2 Omicron sublineages](#), Park YJ, Pinto D et al. *Science*. 2022 Nov 11;378(6620):619-627.  
[Residue forces and nucleotide state jointly regulate F-actin structure](#), Reynolds MJ, Hachicho C et al. *Nature*. 2022 Nov 10;611(7935):380-386.  
[Bestrophen-2 and glutamine synthetase form a complex for glutamate release](#), Owji AP, Yu K et al. *Nature*. 2022 Nov 3;611(7934):180-187.  
[Cryo-EM structures of human m6A writer complexes](#), Su S, Li S et al. *Cell Res*. 2022 Nov;32(11):982-994.  
[Algorithmic design of 3D wireframe RNA polyhedra](#), Elonen A, Natarajan AK et al. *ACS Nano*. 2022 Oct 25;16(10):16608-16616.  
[Previously featured citations...](#)

### Feature Highlight

#### Pipes and Planks

The [PipesAndPlanks](#) tool shows protein helices as "pipes" (cylinders) and strands as "planks" (rectangular boxes), with connectors for the intervening coil. Adjustable settings include pipe radius, plank width, colors, and whether to include arrowheads to show chain N→C directionality (see [image how-to](#)).

[\(More features...\)](#)



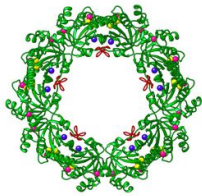
### Gallery Sample

#### Peroxisredoxin Wreath

Peroxisredoxins are enzymes that help cells cope with stressors such as high levels of reactive oxygen species. The image shows a decameric peroxiredoxin from human red blood cells (Protein Data Bank entry [1omv](#)), styled as a holiday wreath.

See also the [RBVI holiday card gallery](#).

[\(More samples...\)](#)



### Chimera Search

Go

Google™ Search

### News

**September 27, 2022**

**Website downtime:** The RBVI website (Chimera, ChimeraX, etc.) and RBVI-hosted web services will be down for maintenance from Tue, Sep 27 9pm PDT, through Wed, possibly extending to Thu, Sep 29 5pm PDT.

**December 20, 2021**

The RBVI wishes you a safe and happy holiday season! See our [2021 card](#) and the [gallery of previous cards](#) back to 1985.

**December 17, 2021**

Chimera production release 1.16 is now available. This will be the last release to support Windows 7. See the [release notes](#) for what's new.

[Previous news...](#)

### Upcoming Events



# Protein preparation

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

*Download Files -> PDB Format*

The screenshot shows the RCSB PDB website interface. At the top, there's a navigation bar with links like 'RCSB PDB', 'Deposit', 'Search', 'Visualize', 'Analyze', 'Download', 'Learn', 'More', 'Documentation', and 'Careers'. Below this is a search bar with the text 'Enter search term(s), Entry ID(s), or sequence'. The main content area displays the entry page for PDB ID 1KE7. On the left, there's a 3D ribbon diagram of the protein structure. On the right, there's a 'Download Files' dropdown menu that is open, showing various file formats. The 'PDB Format' option is selected and highlighted by a red box. Other options include 'FASTA Sequence', 'PDBx/mmCIF Format', 'PDBx/mmCIF Format (gz)', 'PDB Format', 'PDB Format (gz)', 'PDBML/XML Format (gz)', 'Structure Factors (CIF)', and 'Structure Factors (CIF - gz)'. Below the dropdown menu, there's a 'wwPDB Validation' section with a 'Validation Full PDF' link highlighted by a red box. The page also includes a 'Biological Assembly 1' section with a 3D view and a 'Global Symmetry' section.





# 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*

The screenshot displays the RCSB PDB website interface. At the top, the navigation bar includes links for Deposit, Search, Visualize, Analyze, Download, Learn, More, Documentation, and Careers. The main header shows the PDB logo and statistics: 198,528 Structures from the PDB and 1,000,361 Computed Structure Models (CSM). A search bar is present with the text 'Enter search terms, Entry ID(s), or sequence'. Below the header, the 'Structure Summary' tab is selected for entry 1KE7. The entry title is 'CYCLIN-DEPENDENT KINASE 2 (CDK2) COMPLEXED WITH DIHYDRO-2-BENZOTHIEN-5-YLAMINO[METHYLENE]-5-(1-2H-INDOL-2-ONE)'. The PDB DOI is 10.2210/pdb1KE7/pdb. The classification is TRANSFERASE, the organism is Homo sapiens, and the expression system is Spodoptera frugiperda. The deposition date is 2001-11-14 and the release date is 2002-05-14. The experimental data snapshot shows the method as X-RAY DIFFRACTION, resolution as 2.00 Å, R-value free as 0.235, and R-value work as 0.191. The wwPDB Validation section shows a 'Good' status for the Biological Assembly 1 (CIF - gz) and Biological Assembly 1 (PDB - gz). The 'Download Files' dropdown menu is open, showing options for PDBx/mmCIF Format, PDB Format, PDB ML/XML Format, Structure Factors (CIF), and Validation XML. The 'FASTA Sequence' option is highlighted.



# 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)

<https://www.rcsb.org/ligand/LS3>

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

PDB-101 PDB TM Database RCSB PDB

LS3

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

Find entries where: LS3  
is present as a standalone ligand in 1 entries

Find related ligands:  
Similar Ligands (Stereospecific)  
Similar Ligands (including Stereoisomers)  
Similar Ligands (Quick Screen)  
Similar Ligands (Substructure Stereospecific)  
Similar Ligands (Substructure including Stereoisomers)

Small Molecules

Ligands 1 Unique

ID	Chains	Name / Formula / InChI Key	2D Diagram	3D Interactions
LS3	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 <sub>20</sub> H <sub>15</sub> N <sub>3</sub> O <sub>4</sub> S FTQYGLRLRXBPT-IDUWFGFVSA-N		 Ligand Interaction

Download Ideal Coordinates CCD File

Download Instance Coordinates

Chemical Component Summary

Name	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
Identifiers	3-[[[2,2-dioxo-1,3-dihydro-2-benzothiophen-5-yl]amino]methylene]-5-(1,3-oxazol-5-yl)-1H-indol-2-one
Formula	C <sub>20</sub> H <sub>15</sub> N <sub>3</sub> O <sub>4</sub> S
Molecular Weight	393.42
Type	NON-POLYMER
Isomeric SMILES	<chem>O=C1C(=O)C(=O)C(=O)C1=C(C2=CC3=C(C=C2)C(=O)N3)C4=CC=CC=C4</chem>
InChI	InChI=1S/C20H15N3O4S/g24-20-17/22-15-3-1-13-9-28(25,26)10-14(13)/5-15/16-6-12/2-4-18(16)/23-20/19-8-21-11-27-19/h1-8,11,22H,9-10H2,(H,23,24)/b17-7

Chemical Details	
Formal Charge	0
Atom Count	43
Chiral Atom Count	0
Bond Count	47
Aromatic Bond Count	16

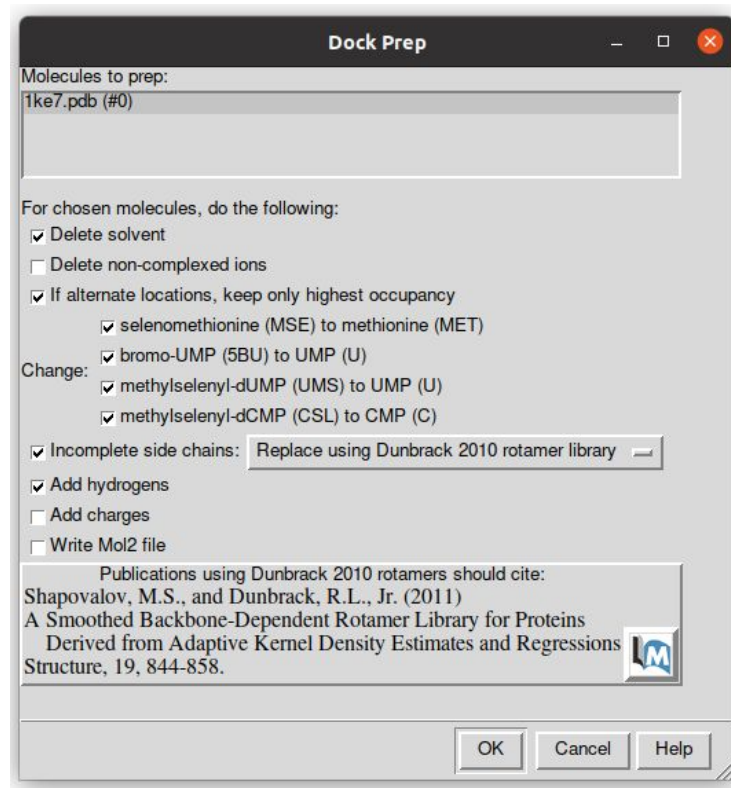


# Protein preparation

## 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)*

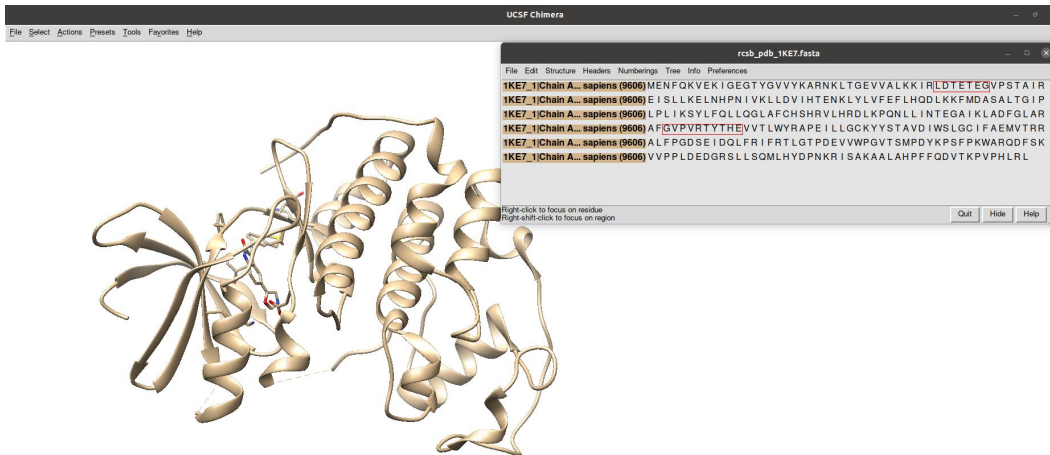
- c. add Hs to selected model

▼ 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***



**Model Loops / Refine Structure**

☐ active region  
☐ Chimera selection region  
Model/remodel: ☐ non-terminal missing structure ☐ all missing structure

1ke7.pdb (#0)

Allow this many residues adjacent to missing regions to move: 1

Number of models to generate: 19

Loop modeling protocol: standard

Run Modeller using: web service

Modeller license key: \*\*\*\*\*

Temporary folder location (optional): Browse

Publications using Modeller results should cite:  
A. Sali and T. L. Blundell.  
Comparative protein modelling by satisfaction of spatial restraints.  
J. Mol. Biol. 234, 779-815, 1993.

OK Apply Modeller Home Page Close Help



# 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*

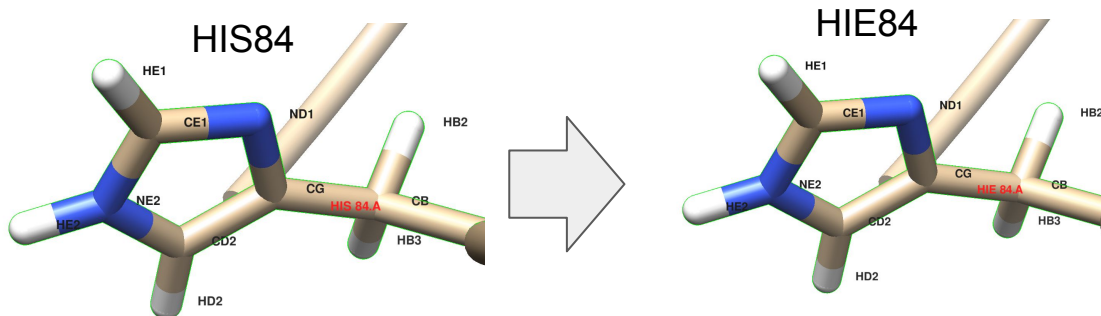
Put in the Command line:

```
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

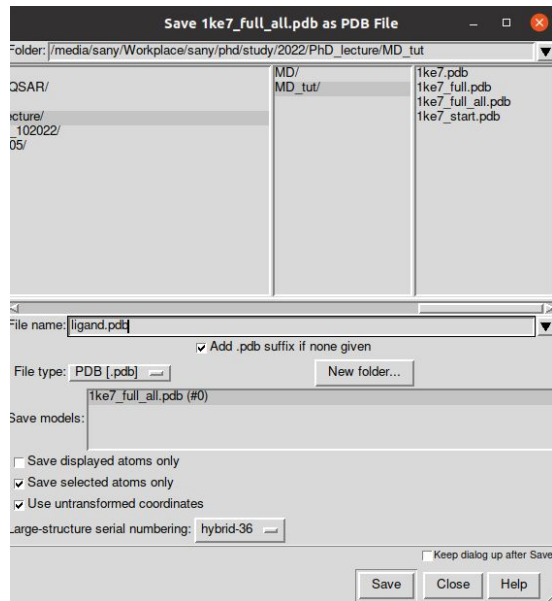
4. Open Fasta and PDB in **Chimera**
  - e. Save protein only (plus ions, metals) and ligand only in *pdb* formats separately

*Select -> Residue -> Ligand\_id*

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

*Select -> Invert (all models)*

*File -> Save PDB -> Save selected only  
protein\_prepared.pdb*





# Ligand preparation

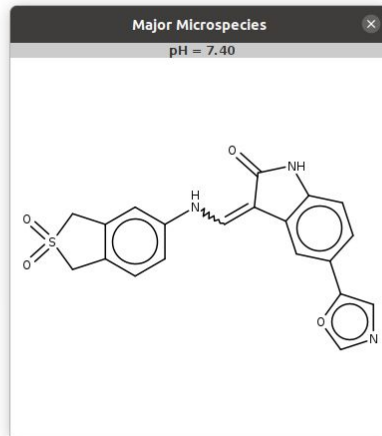
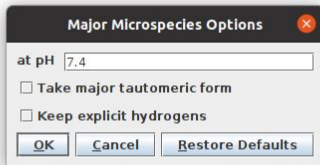
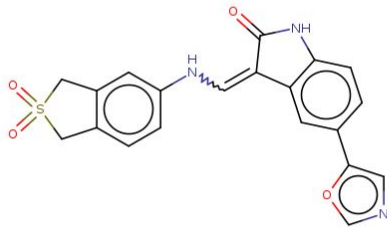
5. Save protonated smiles (pH 7.4) by Marvin into *ligand\_74.smi*

*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

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_74.smi -o ligand.mol
```



# Protein Ligand preparation

Input Files for MD:

## ***protein\_prepared.pdb***

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

## ***ligand.mol***

- *protonated at 7.4 pH*
- *added all hydrogens*

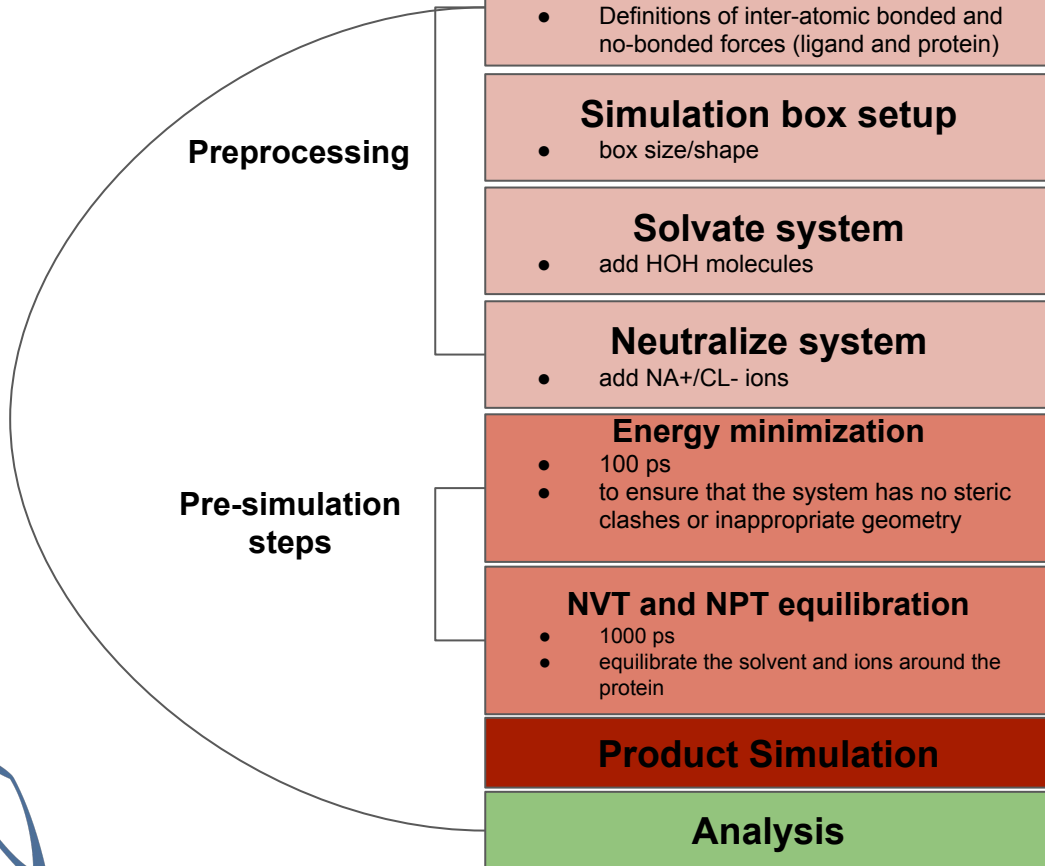
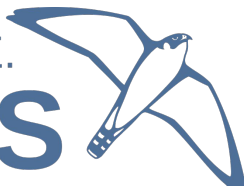


# Classical Molecular Dynamics

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

FAST. FLEXIBLE. FREE.

**GROMACS**





# 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.

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

**New files:**

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



# 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"`



# Automation tools for ligand topology

AMBER	<a href="#"><u>Antechamber</u></a> <a href="#"><u>acpype</u></a>	Parametrizes molecules using GAFF A Python interface to Antechamber, writes GROMACS topologies
CHARMM	<a href="#"><u>CGenFF</u></a>	The official CHARMM General Force Field server
GROMOS87/ GROMOS96	<a href="#"><u>PRODRG 2.5</u></a> <a href="#"><u>ATB</u></a>	An automated server for topology generation A newer server for topology generation, uses GROMOS96 54A7
OPLS-AA	<a href="#"><u>Topolbuild</u></a> <a href="#"><u>TopolGen</u></a> <a href="#"><u>LigParGen</u></a>	Converts a Tripos .mol2 file into a topology A Perl script to convert an all-atom .pdb file to a topology 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<sup>-</sup> or NA<sup>+</sup> to neutralize our protein:

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

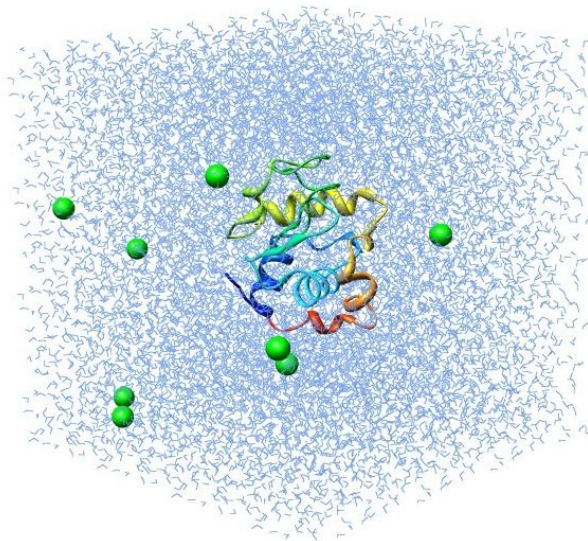
```
gmx genion -s ions.tpr -o protein_solv_ions.gro -p topol.top -pname NA -nname CL -neutral
```

## 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*

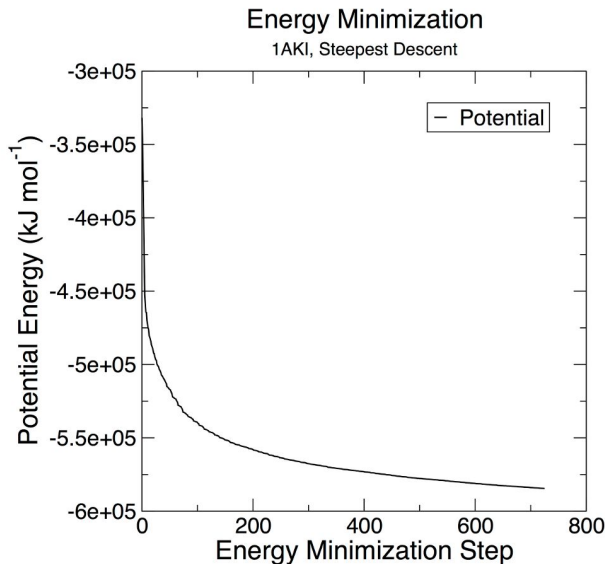


# Energy minimization

## 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"
```







# 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

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

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

$$p \cdot V = N \cdot k_B \cdot T$$

$$p = \rho k_B T$$

$$E = \frac{3}{2} N k_B T$$



# 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

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

- Berendsen thermostat, Anderson thermostat



# NVT equilibration

## to run NVT equilibration:

```
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"
```

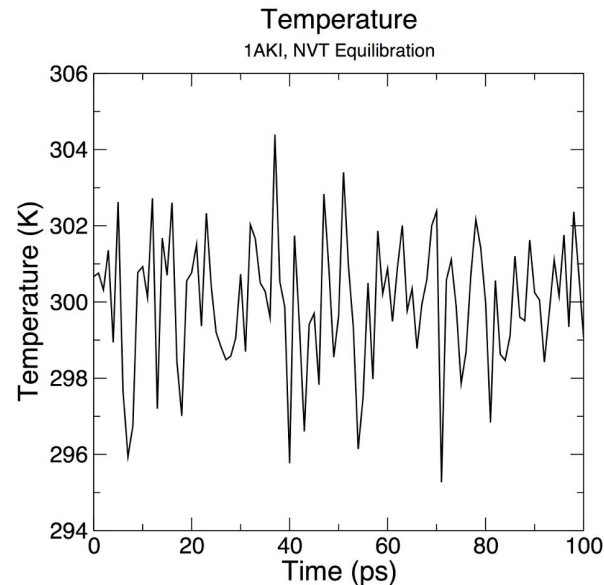
### New files:

*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



# 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 equilibration:

```
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

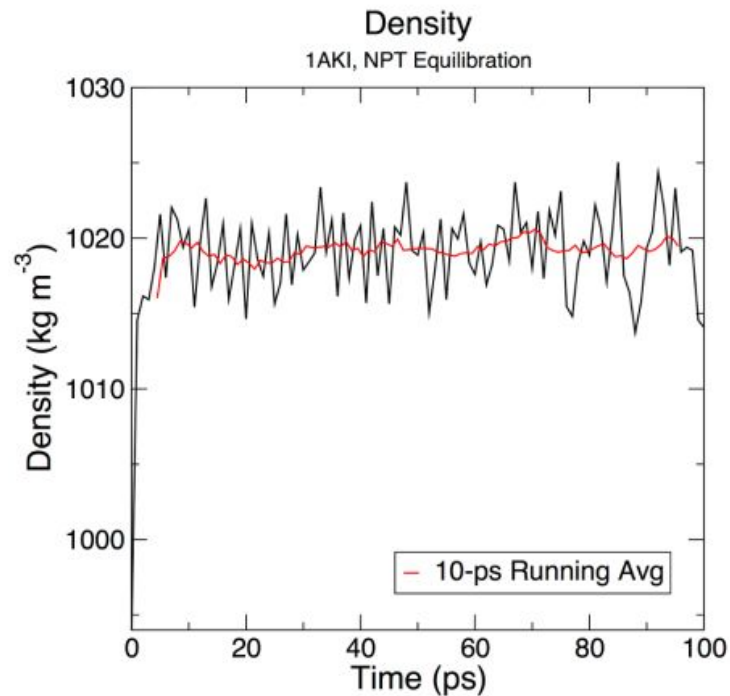
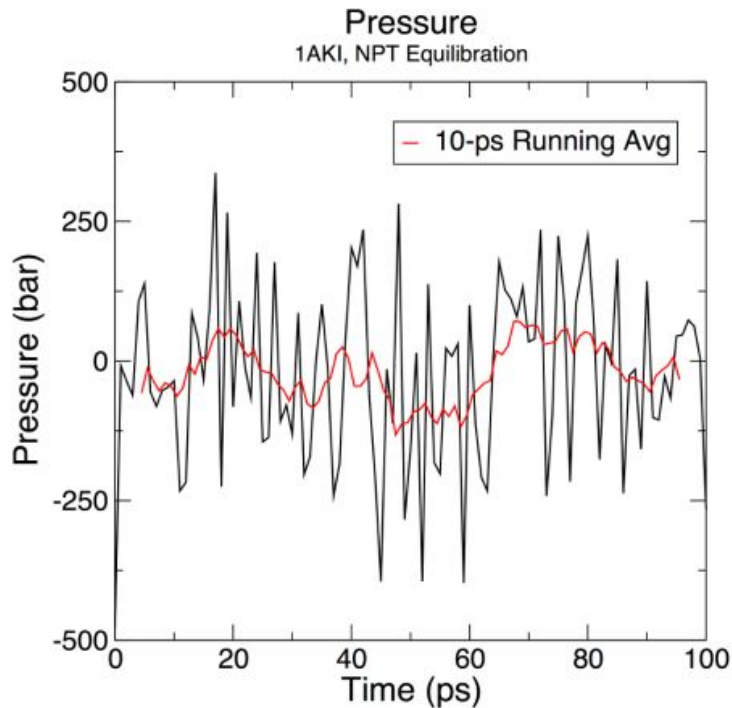
*nvt.edr*: Binary energy file

*nvt.trr*: Binary full-precision trajectory

*nvt.gro*: NVT-minimized structure



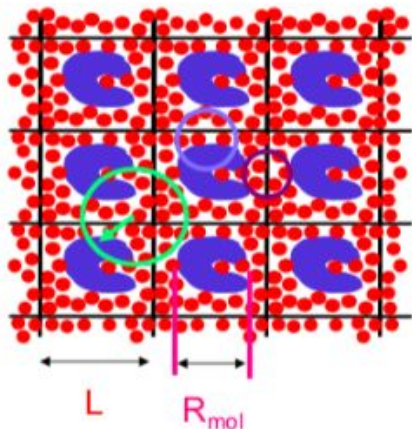
# NPT equilibration



In our tool we run 1000 ps NPT equilibration by default



# Periodic Boundary Conditions



Required  
(no atom sees another one twice):

$$L > 2R_c \quad R_c < \frac{1}{2}L$$

Preferred  
(protein does not see a copy of itself)

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

Even better  
(no solvent sees two proteins)

$$L > 2R_c + R_{mol} \quad R_c < \frac{1}{2}(L - R_{mol})$$

< number of molecules in the box >

• Cubic:

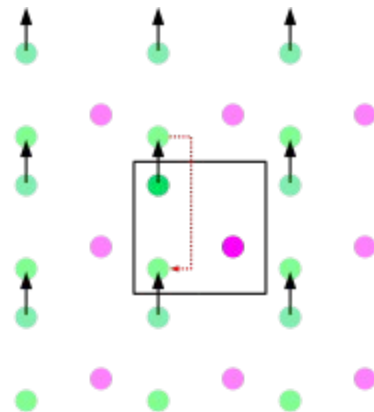


Rectangular:



**Periodic boundary conditions (PBCs)** are a set of boundary conditions which are often chosen for approximating a large (infinite) system by using a small part called a unit cell.

- 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.





# 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

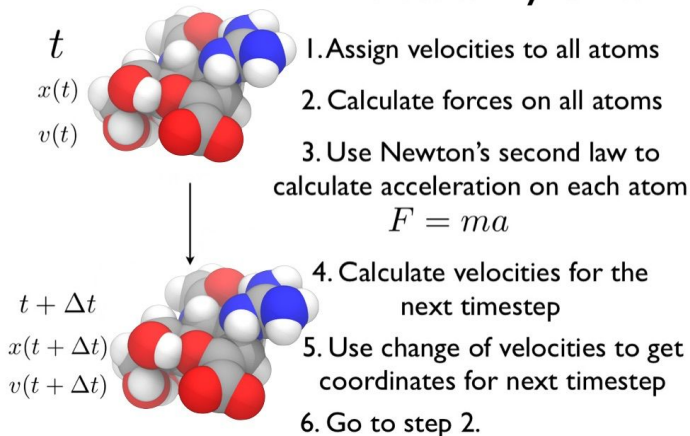
**md\_0\_1.cpt**: portable checkpoint file. The complete state of the simulation is stored in the checkpoint file, including extended thermostat/barostat variables, random number states and NMR time averaged data

**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

## Molecular Dynamics





# 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"
```



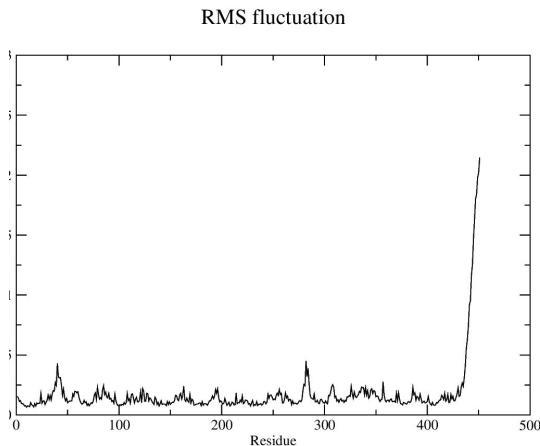
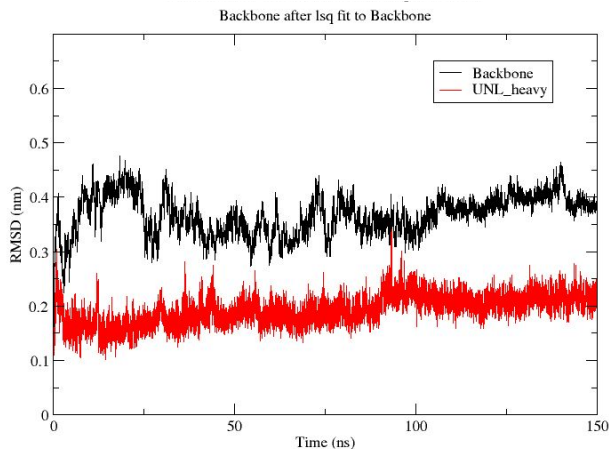
# Analysis of calculated MD simulation

## RMSD:

```
gmx rms -s md_0_1.tpr -f md_0_1_noPBC.xtc -o rmsd.xvg -tu ns
```

## RMSF:

```
gmx rmsf -s md_0_1.tpr -f md_0_1_noPBC.xtc -o rmsf.xvg -oq rmsf.pdb -res
```



xmgrace



## 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.xvg** - rmsd of Backbone of protein

**rmsd\_UNL.xvg** - rmsd of the heavy atoms of ligand

**gyrate.xvg** - radius of gyration of the protein

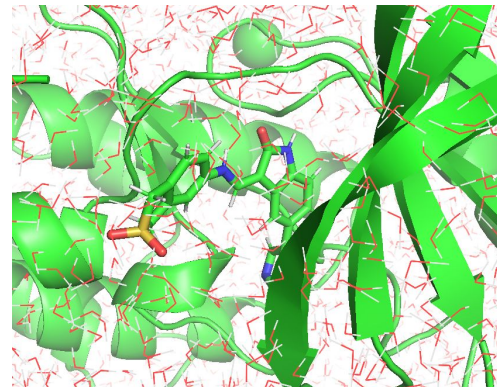
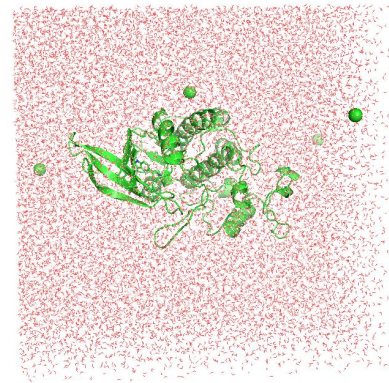
**rmsf.xvg** - root mean square fluctuation of each amino acids





## Check your own MD trajectory

*frame.pdb* - a frame from the trajectory to provide topology  
*md\_fit.xtc* - your fitted MD trajectory





# Analysis of calculated MD simulation

*working dir: /home/user/MD\_tutorial*

**Run on it4i cluster:**

**Visualize plots by xmgrace:**

*module load grace*

*cd ligand\_01ns/*

*for i in \*.xvg; do gracebat -hdevice PNG \$i;done*

```
(base) [antkonen@login4.karolinska.se ~]$ ls
01_complex_preparation_md.pbs.e1627833  en.gro          #index.ndx.2#  ligand.mol2      mdout.mdp        npt.log         posre.itp        rmsd_UNL_xtal.xvg  sqn.out
01_complex_preparation_md.pbs.e1627833  en.log          #index.ndx.3#  ligand.prntop    md_out_noj_noPBC.xtc  npt.mdp        posre_ligand.itp  rmsd_UNL.xvg      sqn.pdb
02_pbsa.pbs.e1627862                    en.tpr          #index.ndx.4#  ligand.top       md_out.tpr        npt.tpr         potential.png      rmsd_xtal.png     temperature.png
02_pbsa.pbs.e1627862                    en.trr          #index.ndx.5#  LIG.prntop       md_out.xtc        npt.trr         pressure.xvg       rmsd_xtal.xvg     temperature.xvg
COMPACT_MMXSA_RESULTS.mmxsa             FINAL_RESULTS_MMPBSA.dat  ions.ndp       md_centermolnoPBC.xtc  md_short_forcheck.xtc  nvt.cpt        pressure.png      rmsd.xvg          tleap.in
complex.gro                             frame.pdb       ions.tpr       md_fit.xtc        mltun.mdp          nvt.edr         pressure.xvg      rmsf.pdb          tnp.gro
COM.prntop                             gmx_MMPBSA.log  ligand.frcmod  md_fit.mdp        mmbsa.in           nvt.gro         protein.gro       rmsf.png          topol.top
COM_traj_0.xtc                         gyrate.png      ligand.gro     md_out.cpt        newbox.gro          nvt.log         REC.prntop       rmsf.xvg          #topol.top.1#
density.png                             gyrate.xvg      ligand.inpcrd  md_out.edr        npt.cpt             nvt.mdp        rmsd.png          solv.gro           #topol.top.2#
density.xvg                             index.ndx       ligand.itp     md_out.gro        npt.edr             nvt.tpr        rmsd_UNL.png      solv_ions.gro     sqn.in
en.edr                                 #index.ndx.1#  ligand.lib     md_out.log        npt.gro             nvt.trr        rmsd_UNL_xtal.png  sqn.in
```

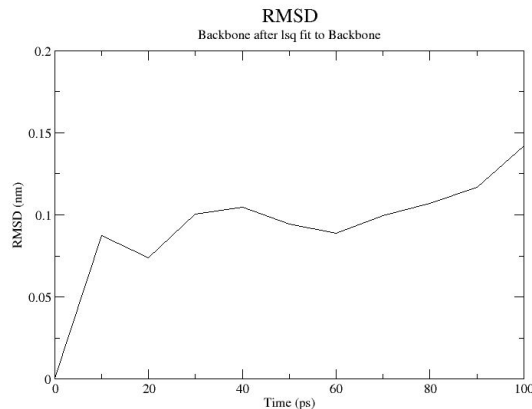
**Run analysis on your own computer:**

*scp -I path/to/your/private/key [dd-23-13-XX@login1.karolinska.it4i.cz](mailto:dd-23-13-XX@login1.karolinska.it4i.cz):md\_tutorial/ligand\_01ns/{frame.pdb,md\_fit.xtc, \*.png} .*

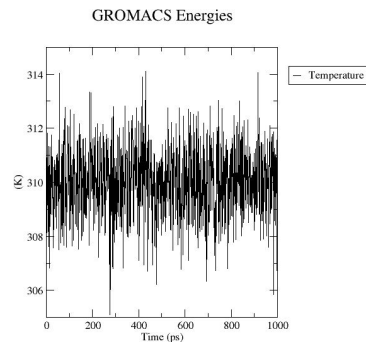
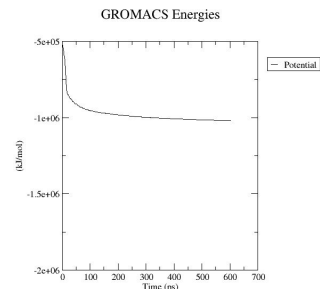
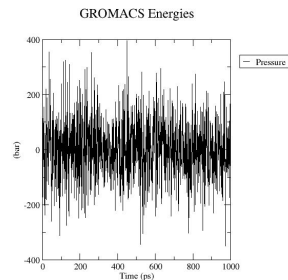
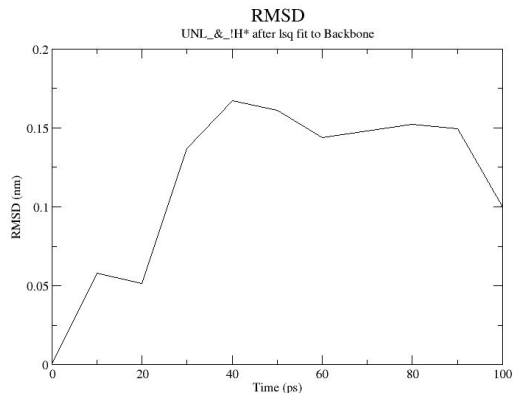


# Analysis of the calculated MD simulation

```
# This file was created Mon Nov 28 18:33:21 2022
# Created by:
#   :~: GROMACS - gmh rms, 2021.4-EasyBuild-4.5.1-PLUMED-2.7.3 (-:
#
# Executable: /apps/all/GROMACS/2021.4-foss-2020b-PLUMED-2.7.3/bin/gmh
# Data prefix: /apps/all/GROMACS/2021.4-foss-2020b-PLUMED-2.7.3
# Working dir: /mnt/proj1/dd-22-84/MD_tutorial/ligand_bins
# Command line:
#   gmh rms -s md_out.tpr -f md_fit.xtc -o rmsd.xvg -n index.ndx -tu ps
# gmh 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"
0.000000 0.0004955
10.000000 0.0872569
20.000000 0.0738152
30.000000 0.1004452
40.000000 0.1048534
50.000000 0.0945487
60.000000 0.0887801
70.000000 0.0993429
80.000000 0.1069565
90.000000 0.1168087
100.000000 0.1428962
```



```
# This file was created Mon Nov 28 18:33:21 2022
# Created by:
#   :~: GROMACS - gmh rms, 2021.4-EasyBuild-4.5.1-PLUMED-2.7.3 (-:
#
# Executable: /apps/all/GROMACS/2021.4-foss-2020b-PLUMED-2.7.3/bin/gmh
# Data prefix: /apps/all/GROMACS/2021.4-foss-2020b-PLUMED-2.7.3
# Working dir: /mnt/proj1/dd-22-84/MD_tutorial/ligand_bins
# Command line:
#   gmh rms -s md_out.tpr -f md_fit.xtc -o rmsd_UNL.xvg -n index.ndx -tu ps
# gmh 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_&_IH* after lsq fit to Backbone"
0.000000 0.0005219
10.000000 0.0578194
20.000000 0.0515443
30.000000 0.1366709
40.000000 0.1673483
50.000000 0.1613055
60.000000 0.1436562
70.000000 0.1480425
80.000000 0.1522363
90.000000 0.1496438
100.000000 0.0998629
```





# MMPBSA / MMGBSA

End-state free energy calculations  
with GROMACS files



GNU nano 2.3.1

File: md-scripts/scripts/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=25, verbose=2, PBRadii=3,
startframe=1, interval=1, verbose=2, PBRadii=3,
interaction_entropy=1, ie_segment=25, temperature=310
/
&gb
igb=5, saltcon=0.150,
/
&pb
istrng=0.15, fillratio=4.0, radiopt=0, indi=1, exdi=80.0
/
```

- 1) You can change **ie\_segment** from 25 to 100

*nano /mnt/proj2/dd-23-13/md-scripts/scripts/mmpbsa.in*

***working dir: /home/user/md\_tutorial***

- 2) run script for energy calculation:

**qsub -A DD-22-13 -v tpr=md\_out.tpr,xtc=md\_fit.xtc,script\_path=/mnt/proj2/dd-23-13/md-scripts/scripts/,wdir=ligand\_01ns/,NP=11  
/mnt/proj2/dd-23-13/md-scripts/Protein-Ligand\_MD\_PBGBSA/02\_pbsa.pbs**

- 3) After calculation will be finished

**cat ligand\_01ns/FINAL\_RESULTS\_MMPBSA.dat**



For multiple molecules you can run other script to collect MM(P/G)BSA energies from all directories

*wdir=\$(pwd)/md\_tutorial bash 03\_sum\_result\_pbsa.pbs*

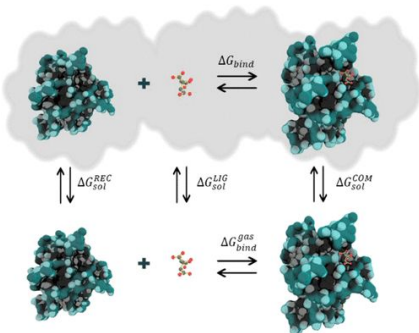
```
Sum_Result_GENERALIZED_BORN.csv  Sum_Result_IE.csv  Sum_Result_POISSON_BOLTZMANN.csv
```

```
(miniconda3)[anikonenko1@login2.karolina dd-23-13]$ cat md_tutorial/Sum_Result_*.  
fname,deltaG,SD  
1ke5_ligand,-30.6897,4.1740  
1ke6_ligand,-23.3582,2.7462  
1ke7_ligand,-19.3363,4.8091  
fname,IE,SD,SEM  
1ke5_ligand,3.8787,2.5746,1.4865  
1ke6_ligand,9.3748,0.0504,0.0291  
1ke7_ligand,12.6552,0.0505,0.0291
```



# 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),

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

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

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

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

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



# What can be done by MD

- to estimate binding affinity of protein-ligand complexes

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)

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

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

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

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

**Total  $G_{\text{Binding}}$  =**

- Gas-phase molecular mechanics energy  $\Delta E_{\text{MM}}$ :**

- changes in the **internal energies**  $\Delta E_{\text{int}}$  (bond, angle, and dihedral energies)
- electrostatic energies**  $\Delta E_{\text{ele}}$
- van der Waals energies**  $\Delta E_{\text{vdW}}$

- sum of the electrostatic solvation energy  $G_{\text{sol}}$**

- The polar contribution** is calculated using either the PB or GB model ( $\Delta G_{\text{PB/GB}}$ ). 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 ( $\text{SASA}$ )

- 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 which





EDISPER	-1407.08	18.20	18.20	5.49	5.49
GGAS	-3184.42	101.56	74.86	30.62	22.57
GSOLV	-2452.28	57.05	52.79	17.20	15.92
TOTAL	-5636.70	116.49	31.95	35.12	9.63

Ligand:					
Energy Component	Average	SD(Prop.)	SD	SEM(Prop.)	SEM
BOND	11.40	2.28	2.28	0.69	0.69
ANGLE	45.46	2.79	2.79	0.84	0.84
DIHED	23.24	2.06	2.06	0.62	0.62
VDWAALS	-3.31	0.62	0.62	0.19	0.19
EEL	95.11	0.99	0.99	0.30	0.30
1-4 VDW	7.97	0.54	0.54	0.16	0.16
1-4 EEL	-225.15	1.58	1.58	0.48	0.48
EPB	-34.82	0.85	0.85	0.26	0.26
ENPOLAR	40.24	0.17	0.17	0.05	0.05
EDISPER	-42.92	0.18	0.18	0.06	0.06

GGAS	-45.28	4.62	3.37	1.39	1.39
GSOLV	-37.50	0.89	0.95	0.27	0.27
TOTAL	-82.78	4.71	2.96	1.42	1.42

Delta (Complex - Receptor - Ligand):					
Energy Component	Average	SD(Prop.)	SD	SEM(Prop.)	SEM
ΔBOND	-0.00	2.08	0.00	0.63	0.00
ΔANGLE	-0.00	2.09	0.00	0.63	0.00
ΔDIHED	0.00	1.45	0.00	0.44	0.00
ΔVDWAALS	-46.46	0.60	2.50	0.18	0.75
ΔEEL	-41.29	0.26	7.77	0.08	2.34
Δ1-4 VDW	-0.00	0.30	0.00	0.09	0.00
Δ1-4 EEL	0.00	1.33	0.00	0.40	0.00
ΔEPB	64.17	0.71	4.26	0.22	1.28
ΔENPOLAR	-32.09	0.07	0.75	0.02	0.23
ΔEDISPER	57.51	0.07	0.85	0.02	0.26

ΔGGAS	-87.75	0.66	7.62	0.20	2.30
ΔGSOLV	89.59	0.72	4.38	0.22	1.32
ΔTOTAL	1.84	0.98	6.93	0.29	2.09

Using Interaction Entropy Approximation:  
ΔG binding = 4.51 +/- 6.93

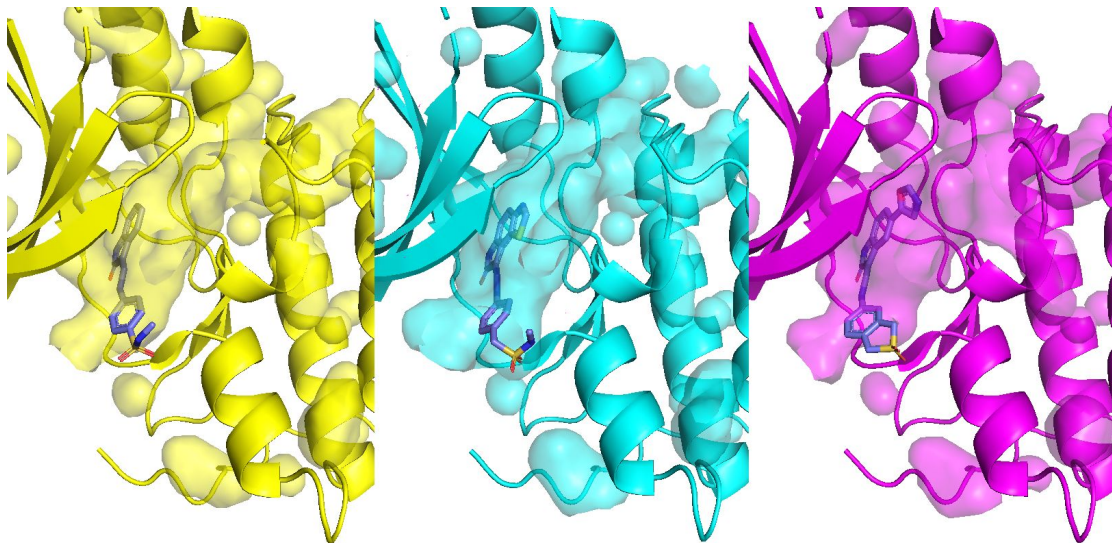
GENERALIZED BORN:  
POISSON BOLTZMANN:

```
(base) [anikonenko1@login4.karolina MD_tutorial]$ grep 'G binding =' ligand_01ns/FINAL_RESULTS_MMPBSA.dat
ΔG binding = -36.09 +/- 4.58
ΔG binding = 4.51 +/- 6.93
```

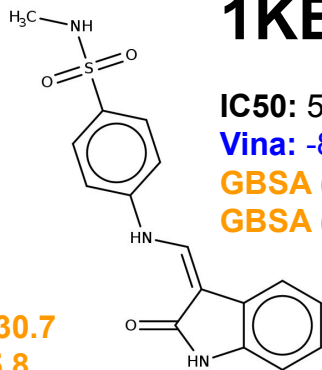
**MMPBSA Energy and MMGBSA Energy cannot be compared within the different methods.** But you can rank your molecules by energies obtained from each method separately.



## CYCLIN-DEPENDENT KINASE 2 (CDK2)



**1KE5**



**IC50:** 5.7 nM

**Vina:** -8.9

**GBSA (0.1ns):** -23.3

**GBSA (1ns):** -21.5

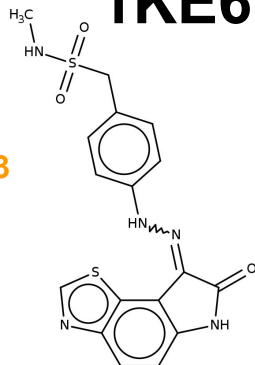
**IC50:** 560 nM

**Vina:** -8.7

**GBSA (0.1 ns):** -30.7

**GBSA (1 ns):** -25.8

**1KE6**



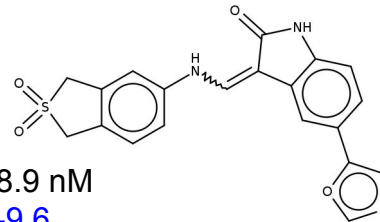
**IC50:** 8.9 nM

**Vina:** -9.6

**GBSA (0.1ns):** -19.3

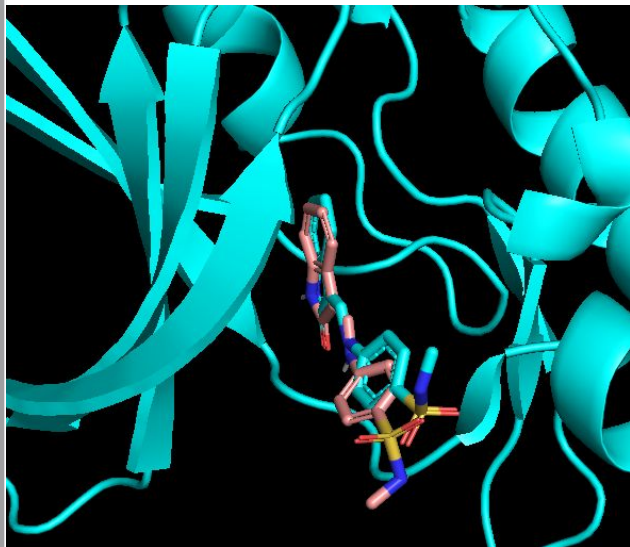
**GBSA (1ns):** -27.1

**1KE7**

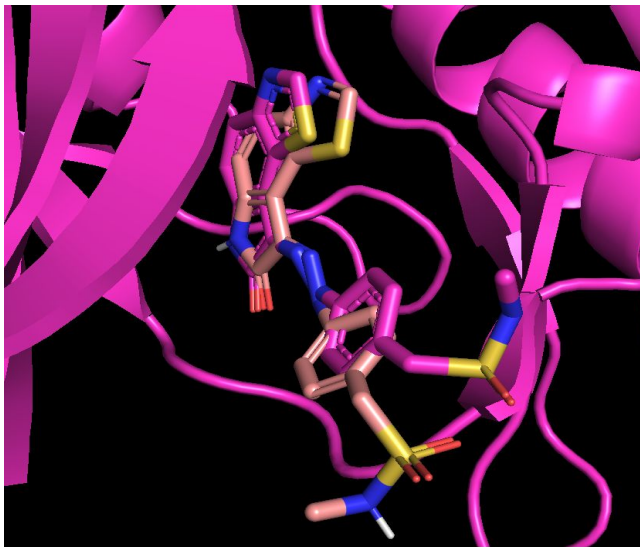




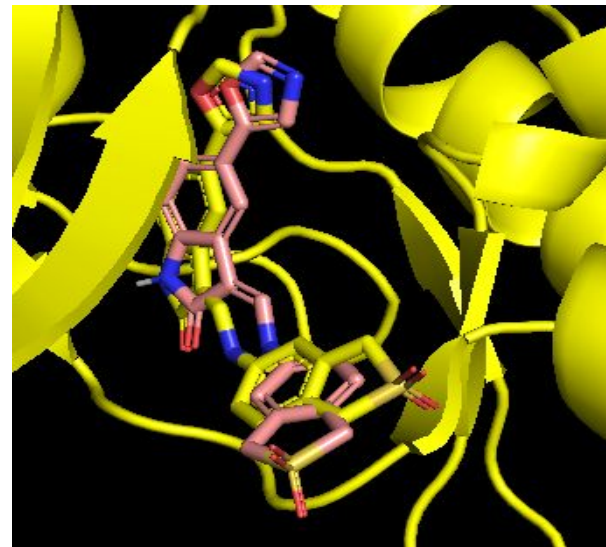
## Docking best pose



**1KE5**



**1KE6**

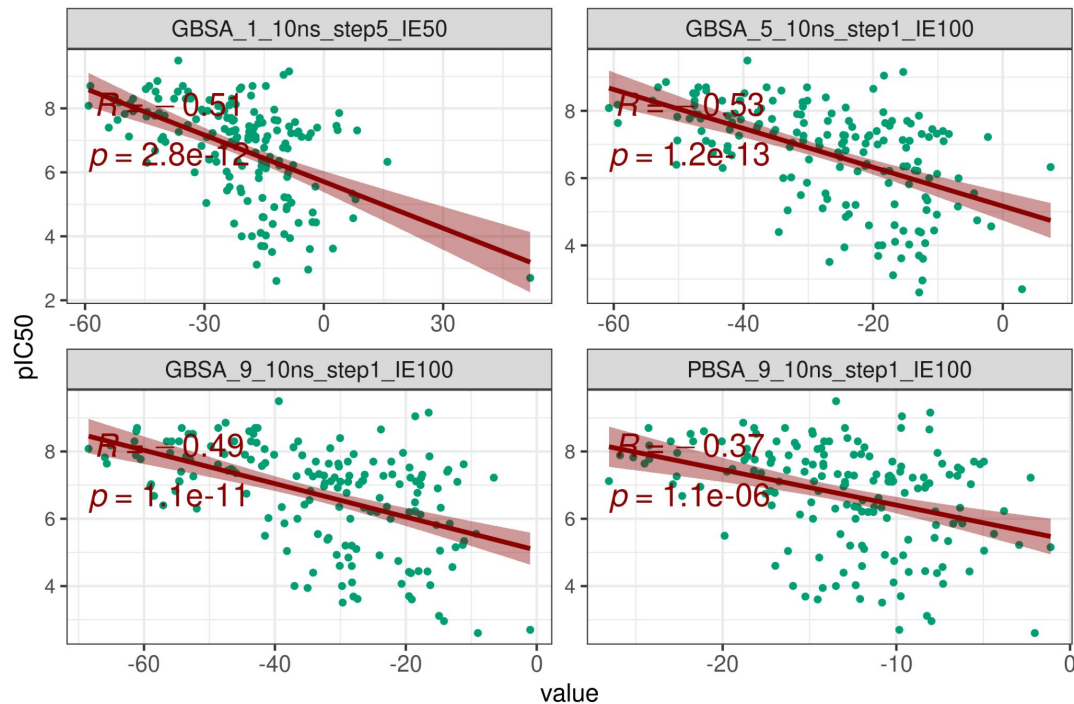


**1KE7**



# Beta Secretase

10ns. 310K. GBSA. Pearson





# **Thank you for your attention!**