High-throughput Molecular Dynamics Workshop

Aleksandra Ivanova
PhD student
Palacký University
Supervisor: **Dr. Pavel Polishchuk**

INSTITUTE OF MOLECULAR AND TRANSLATIONAL MEDICINE

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

RCSB PDB PROTEIN DATA BANK — 198,528 Structures from the PDB; 1,000,361 Computed Structure Models (CSM)

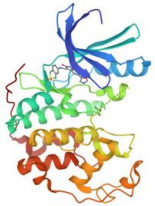
3D Structures | Enter search term(s), Entry ID(s), or sequence | Include CSM

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

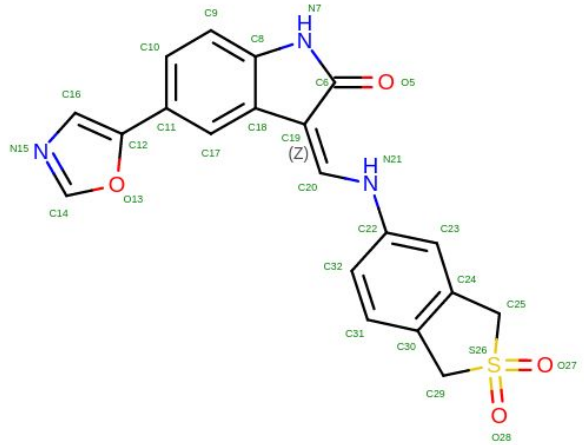
Entity ID: 1

Molecule	Chains	Sequence Length	Organism	Details	Image
Cell division protein kinase 2	A	298	Homo sapiens	**Mutation(s):** 0 **Gene Names:** CDK2, CDK N2 **EC:** 2.7.1.37 (PDB Primary Data), 2.7.1.22 (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):

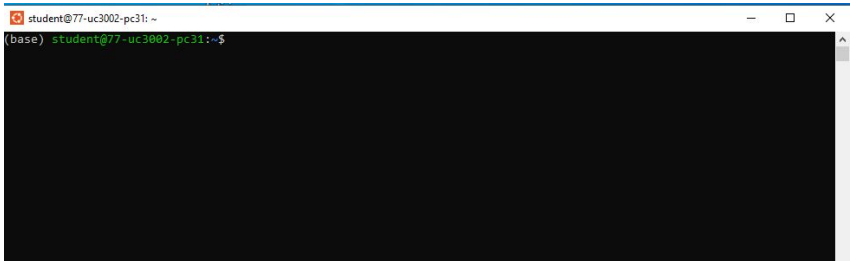
download data from https://www.kfc.upol.cz/8add

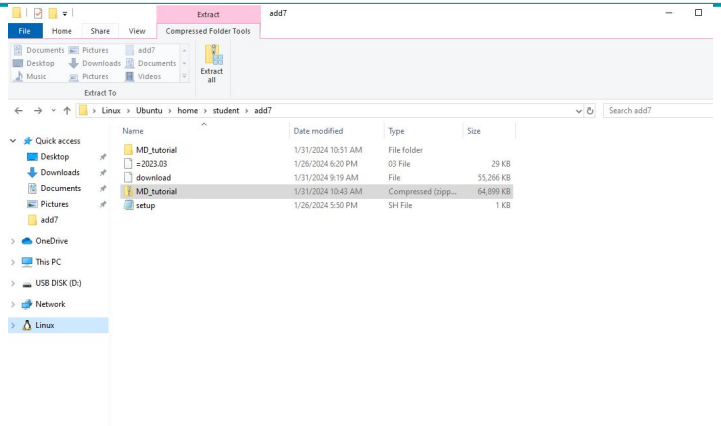
mkdir 8add

*unzip manually under **Ubuntu/home/student/8add** directory*

cd 8add/MD_tutorial/

conda activate md





Run MD simulation (100 ps):

run_md *-p protein_H_HIS.pdb -l ligand.mol --md_time 0.1 --nvt_time 10 --npt_time 10 --ncpu 8 -d mdrun --device gpu*

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
```

```
2024-01-31 09:39:22,455 - distributed.comm.tcp - ERROR - Could not set timeout on TCP stream
Traceback (most recent call last):
  File "/home/student/miniconda3/envs/md/lib/python3.9/site-packages/distributed
    sock.setsockopt(socket.SOL_TCP, TCP_USER_TIMEOUT, timeout * 1000)
OSError: [Errno 92] Protocol not available
2024-01-31 09:39:22,456 - distributed.comm.tcp - ERROR - Could not set timeout o
Traceback (most recent call last):
  File "/home/student/miniconda3/envs/md/lib/python3.9/site-packages/distributed
    sock.setsockopt(socket.SOL_TCP, TCP_USER_TIMEOUT, timeout * 1000)
OSError: [Errno 92] Protocol not available
2024-01-31 09:39:22,480 - distributed.comm.tcp - ERROR - Could not set timeout o
Traceback (most recent call last):
  File "/home/student/miniconda3/envs/md/lib/python3.9/site-packages/distributed
    sock.setsockopt(socket.SOL_TCP, TCP_USER_TIMEOUT, timeout * 1000)
OSError: [Errno 92] Protocol not available
2024-01-31 09:39:22,480 - distributed.comm.tcp - ERROR - Could not set timeout o
Traceback (most recent call last):
  File "/home/student/miniconda3/envs/md/lib/python3.9/site-packages/distributed
    sock.setsockopt(socket.SOL_TCP, TCP_USER_TIMEOUT, timeout * 1000)
OSError: [Errno 92] Protocol not available
2024-01-31 09:39:22,495 - distributed.core - INFO - Connection to tcp://127.0.0.
2024-01-31 09:39:22 - root - INFO: Analysis of md simulation of 1 were successfu
Finished: ['/home/student/add7/MD_tutorial/MD_tutorial/mdrun/md_files/md_run/pro
(md) student@77-uc3002-pc31:~/add7/MD_tutorial/MD_tutorial$
(md) student@77-uc3002-pc31:~/add7/MD_tutorial/MD_tutorial$
```

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

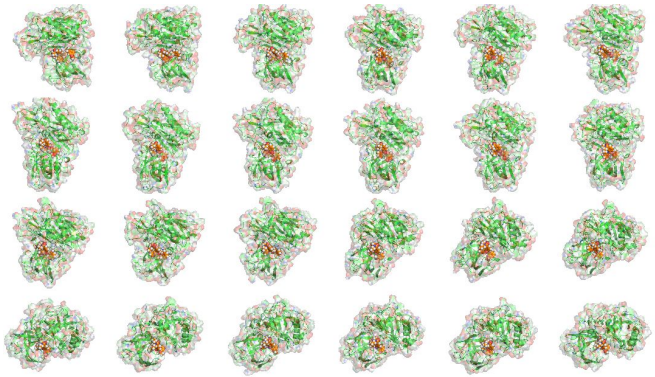
File Edit Format View Help

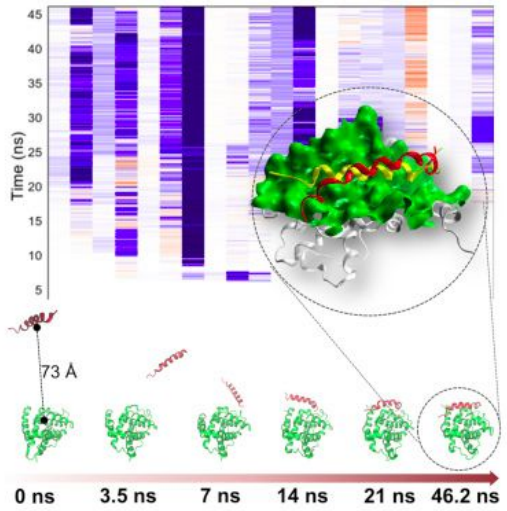
```
2024-01-31 09:23:54 - root - INFO: Namespace(protein='/home/student/add7/MD_tutorial/MD_t
2024-01-31 09:23:54 - root - INFO: Start protein preparation
2024-01-31 09:23:55 - root - INFO: Successfully finished protein preparation

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.**

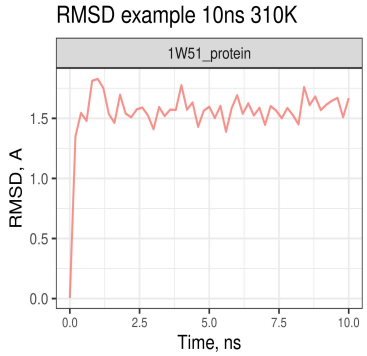
- to explore conformational space
- to explore biological process of molecular recognition
- to investigate ligand pose stability
- to explore protein flexibility
- to estimate binding affinity of protein-ligand (protein-protein) complexes

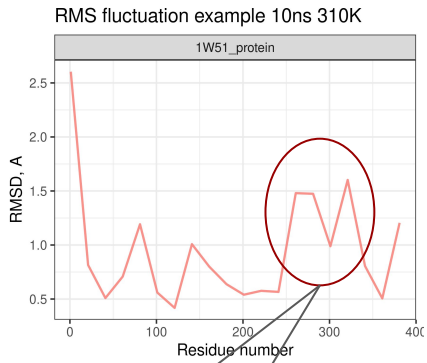




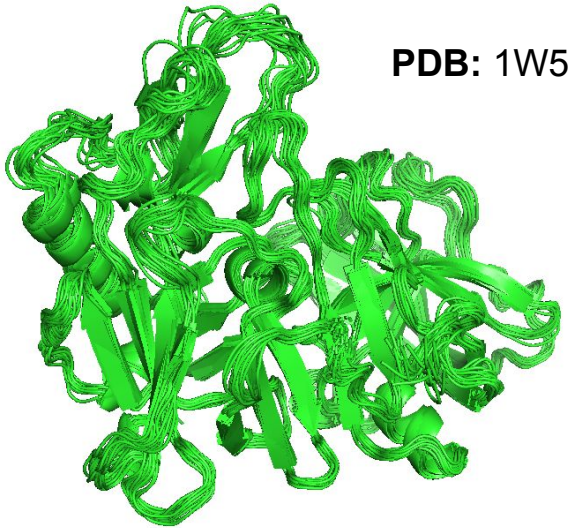
https://doi.org/10.1016/j.str.2017.02.009

What can be done by MD

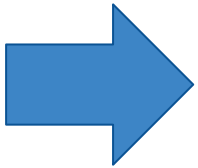




- To explore different conformations of protein
  - To investigate internal-flexibility of protein
  - For practical use we select conformations from the most populated clusters of all conformations

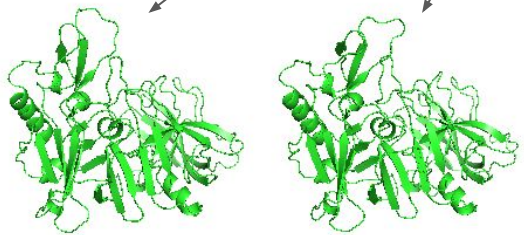


PDB: 1W51



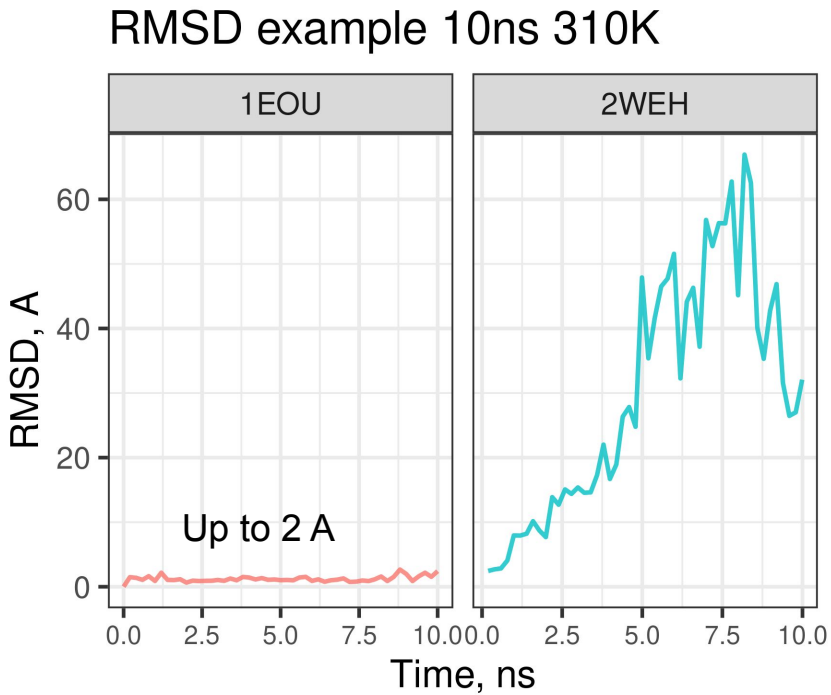
Conformation can be used:

- for further molecular docking study

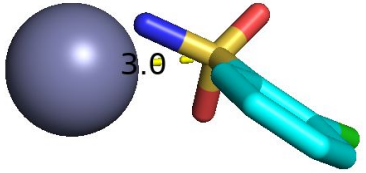


What can be done by MD

- To explore stability of ligand pose

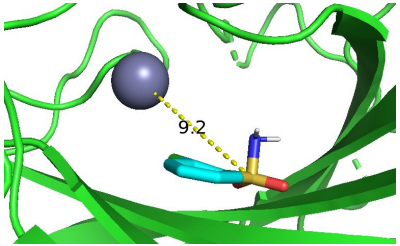


Example of incorrect protonation and pose:

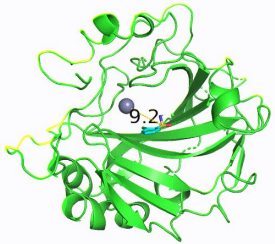


x-ray

PDB: 2WEH



docking

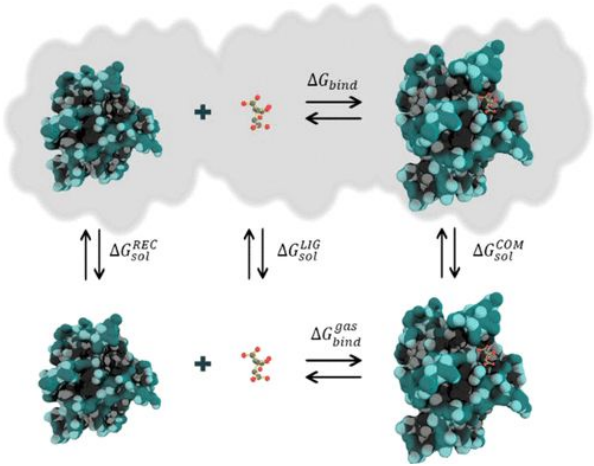


Carbonic Anhydrase

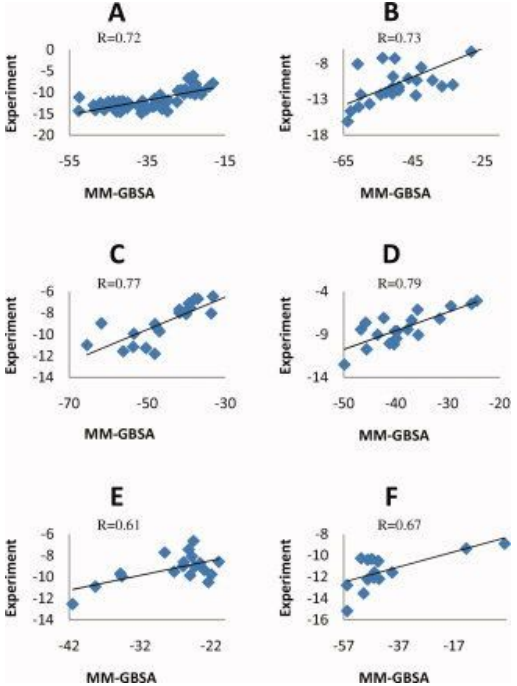
MD 10ns.
Ligand goes out of the site

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.

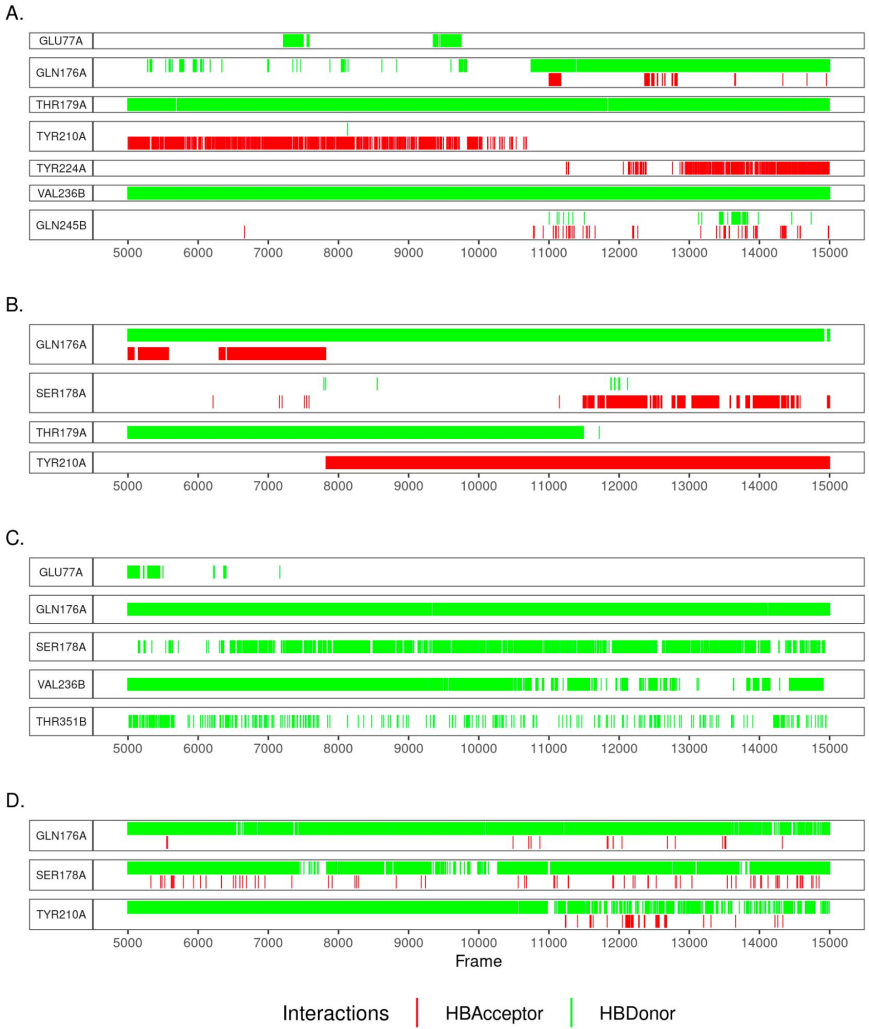
***MMPBSA.py*: An Efficient Program for End-State Free Energy Calculations** Bill R. Miller III, T. Dwight McGee Jr., Jason M. Swails, Nadine Homeyer, Holger Gohlke, and Adrian E. Roitberg Journal of Chemical Theory and Computation **2012** *8* (9), 3314-3321 DOI: 10.1021/ct300418h

What can be done by MD

- to investigate protein-ligand interaction stability

[5]:

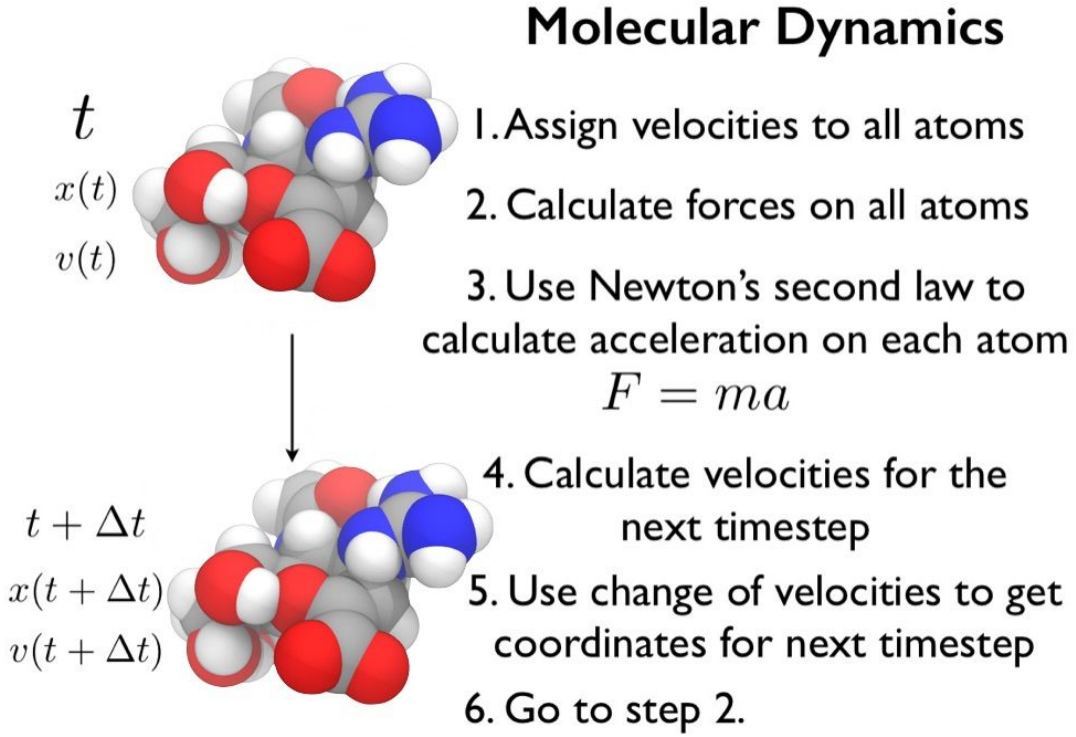
ligand	LIG1.G						
protein	TYR38.A		TYR109.A		THR110.A	TRP125.A	
interaction	Hydrophobic	VdWContact	Hydrophobic	VdWContact	Hydrophobic	Hydrophobic	VdWContact
Frame							
0	False	False	True	False	False	True	False
10	False	False	True	True	False	True	False
20	False	False	True	True	False	True	True
30	True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



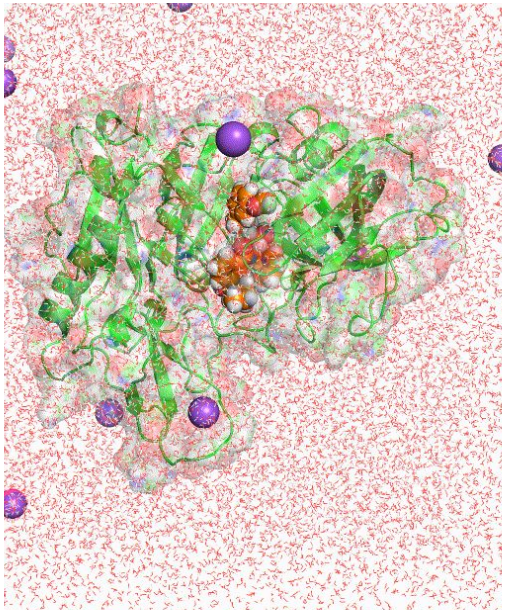
Bouysset, C., Fiorucci, S. ProLIF: a library to encode molecular interactions as fingerprints.
J Cheminform 13, 72 (2021).
https://doi.org/10.1186/s13321-021-00548-6

Molecular Dynamics Simulation Process

Simulation process is based on Newton's second law



Classical Molecular Dynamics



Structure Preparation

Preprocessing

Force-field
- Definitions of inter-atomic bonded and non-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

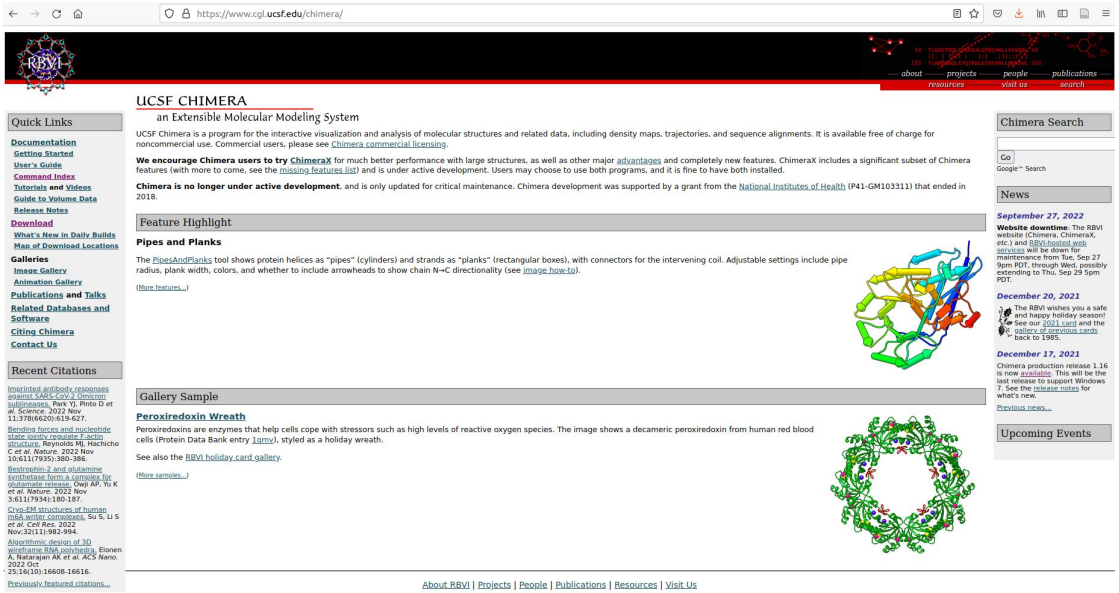
NVT and NPT equilibration
- 1000 ps
- equilibrate the solvent and ions around the protein

Product Simulation

Analysis

FAST. FLEXIBLE. FREE.
GROMACS

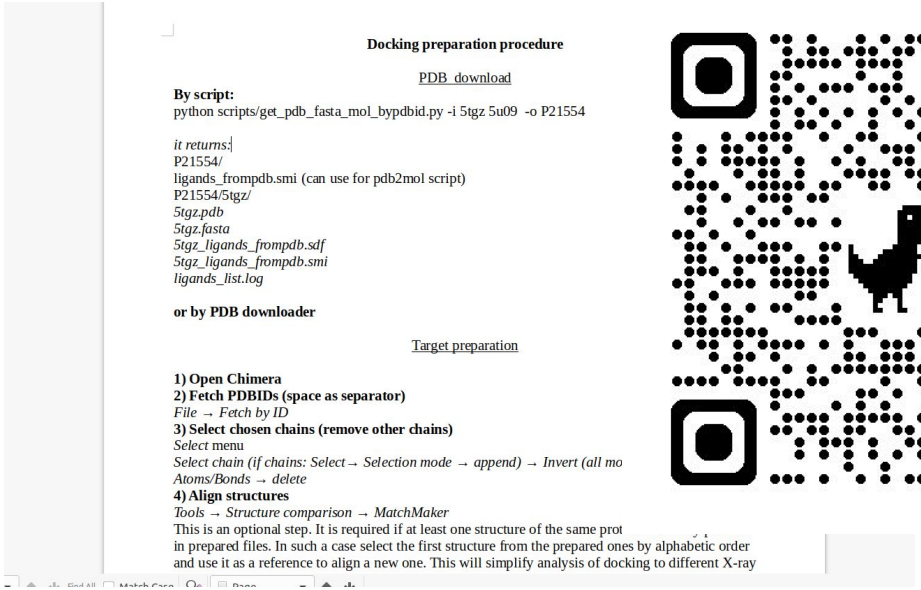
Structure preparation



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

Protein and ligand preparation. In-house scripts

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



Docking preparation procedure

PDB download

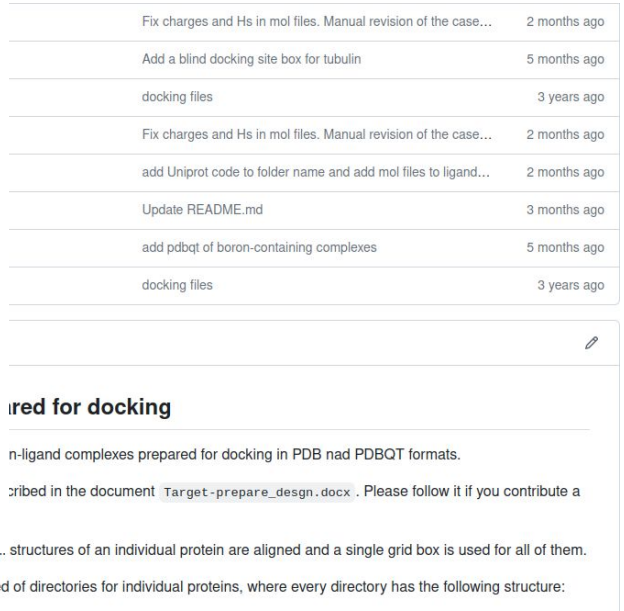
By script:
python scripts/get_pdb_fasta_mol_bypdbid.py -i 5tgz 5u09 -o P21554

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 mo
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 prot
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



Fix charges and Hs in mol files. Manual revision of the case...	2 months ago
Add a blind docking site box for tubulin	5 months ago
docking files	3 years ago
Fix charges and Hs in mol files. Manual revision of the case...	2 months ago
add Uniprot code to folder name and add mol files to ligand...	2 months ago
Update README.md	3 months ago
add pdbqt of boron-containing complexes	5 months ago
docking files	3 years ago

ıred for docking

n-ligand complexes prepared for docking in PDB nad PDBQT formats.

ɔribed in the document `Target-prepare_desgn.docx`. Please follow it if you contribute a

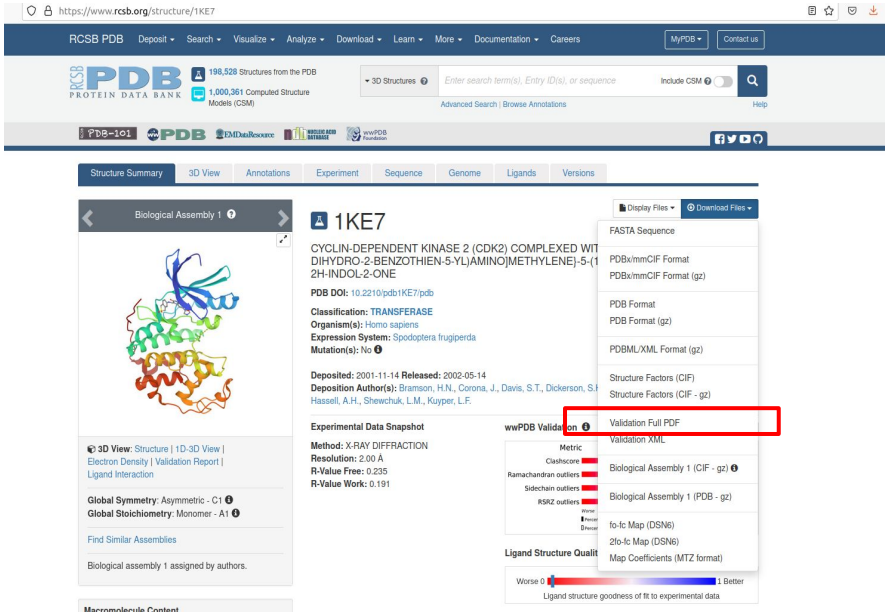
structures of an individual protein are aligned and a single grid box is used for all of them.

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

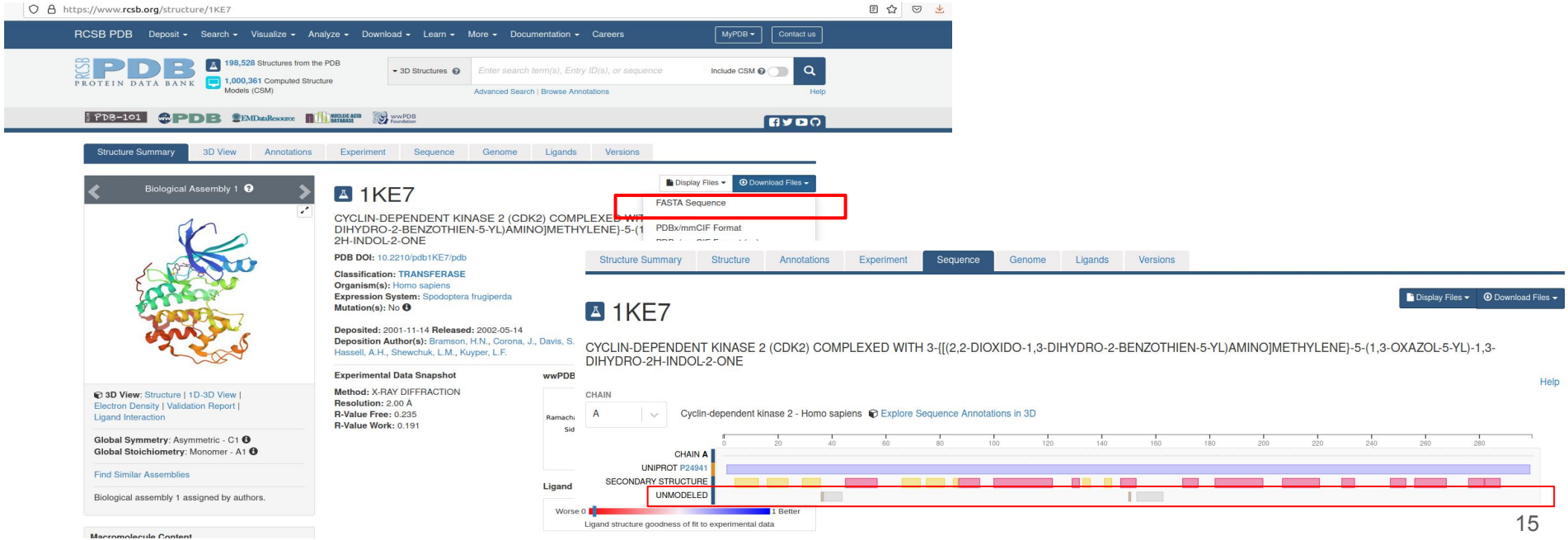
   Download Files -> PDB Format



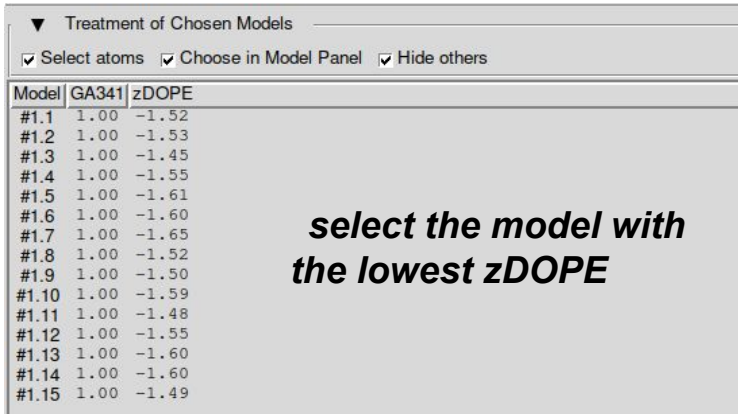
Protein preparation

2. Download sequence from PDB or from UniProt

Download Files -> Fasta Sequence



Protein preparation

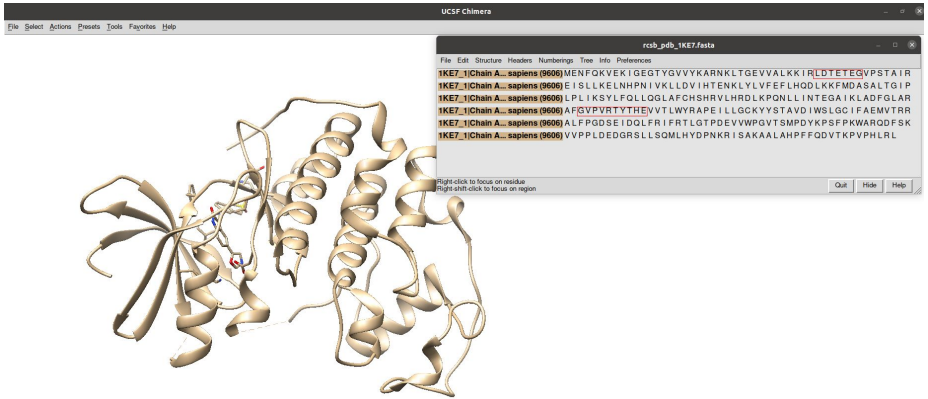


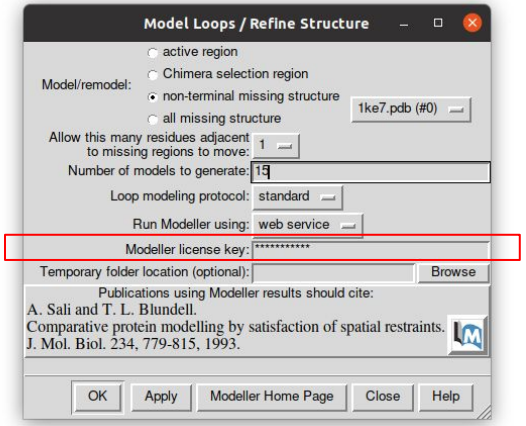
select the model with the lowest zDOPE

3. Open Fasta and PDB in **Chimera**
   a. Fill missing loops by Modeller

Tools -> Sequence -> Sequence

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



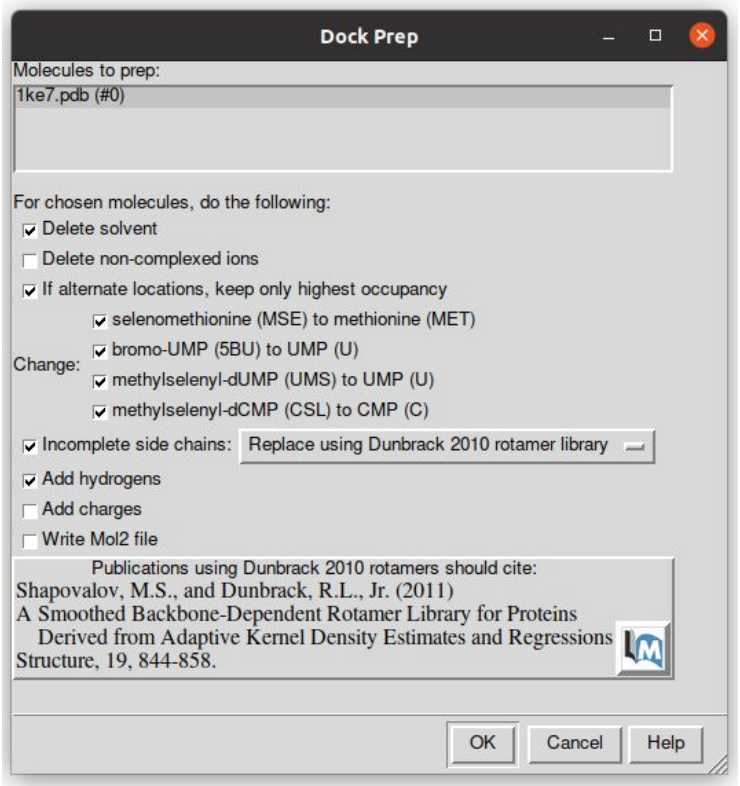


Protein preparation

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

   b. Dock Prepare

      Structure Editing -> Dock Prep



Protein preparation

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

   c. Write *protein, ligands, cofactors* objects into separate files

   Select -> Residue -> choose ligand or cofactor name

   Select -> Structure -> protein

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. Assign non-standard Amber resnames

pdb4amber -i protein_H.pdb -o protein_H_HIS.pdb

- **Histidine (HID, HIE, HIP)**

Can be protonated at different nitrogen atoms in its imidazole ring.

- **Aspartate (ASP, ASH)**

Can be neutral (ASP) or negatively charged (ASH) based on the protonation of the carboxyl group.

- **Glutamate (GLU, GLH)**

Behaves similarly to aspartate with its carboxyl groups.

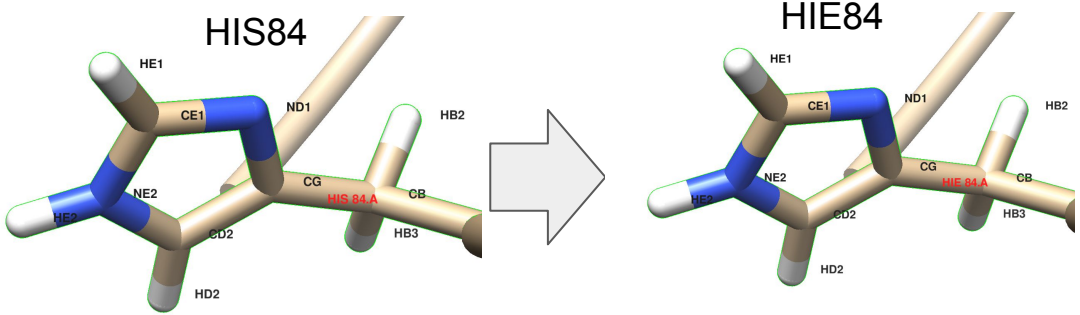
- **Lysine (LYS, LYN)**

Can be protonated or neutral depending on the nitrogen atom in the side chain amine group.

- **Cysteine (CYS, CYX)**

Cysteines involved in disulfide bridges have a special Amber resname (CYX)

Check visually active site



Protein / Ligand preparation

Input Files for MD:

protein_H_HIS.pdb

- *no missing non-terminal atoms/residues*
- *removed non-protein organic molecules*
- *added all hydrogens*
- *set non-standard protonation states of residues*
- *set CYX (Cysteines forming disulfide bridges)*

ligand.mol

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

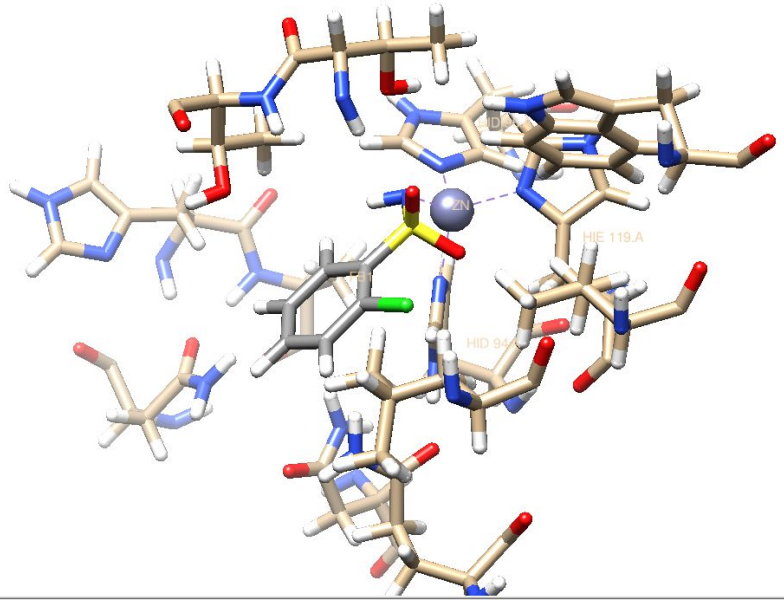
Ligand preparation

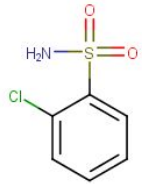
pH-based protonation (pH 7.4)

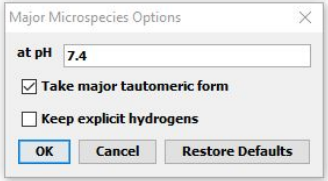
2-CHLOROBENZENESULFONAMIDE

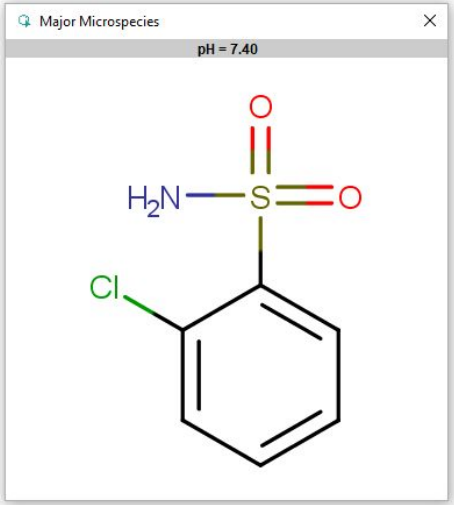
3D-structure-based protonation

Carbonic Anhydrase



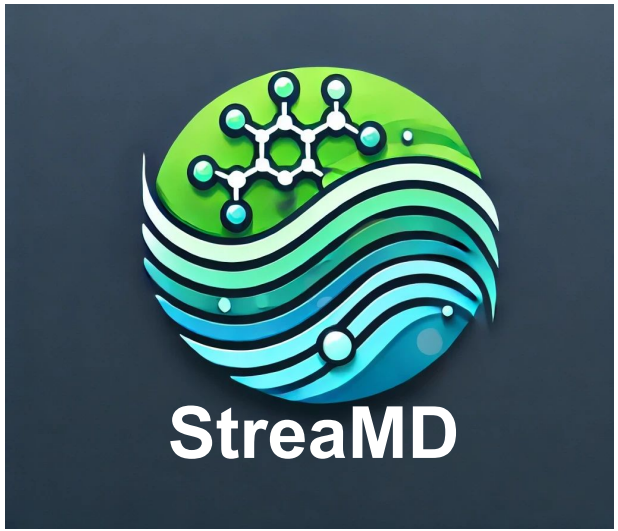






2WEH

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
- 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

run_md -p protein.pdb -l ligands.sdf --md_time 1

INSTITUTE OF MOLECULAR AND TRANSLATIONAL MEDICINE

FAST. FLEXIBLE. FREE. GROMACS

python™

RDKit
Open-Source Cheminformatics and Machine Learning

MD ANALYSIS

Journal of Cheminformatics

Home About Articles Submission Guidelines About The Editors Calls For Papers Submit manuscript

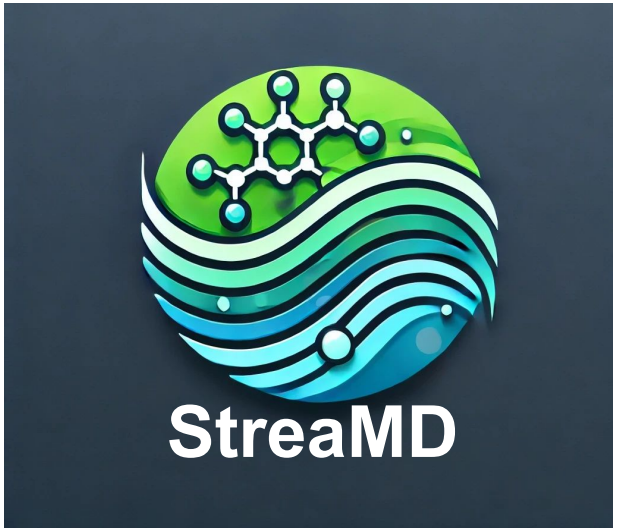
Software | Open access | Published: 05 November 2024

StreaMD: the toolkit for high-throughput molecular dynamics simulations

Aleksandra Ivanova, Olena Mokshyna & Pavel Polishchuk

Journal of Cheminformatics **16**, Article number: 123 (2024) | Cite this article

2072 Accesses | **1** Citations | **6** Altmetric | Metrics



https://github.com/ci-lab-cz/streamd

```
run_md -h
usage: run_md [-h] [-p FILENAME] [-d WDIR] [-l FILENAME] [--cofactor FILENAME] [--clean_previou
              [--topol_itp topol_chainA.itp topol_chainB.itp [topol_chainA.itp topol_chainB.itp
              [--protein_forcefield amber99sb-ildn] [--md_time ns] [--npt_time ps] [--nvt_time
              [--wdir_to_continue DIRNAME [DIRNAME ...]] [--deffnm preffix for md files] [--tpr
              [--ligand_list_file all_ligand_resid.txt] [--ligand_id UNL] [--activate_gaussian
              [--gaussian_exe g09 or /apps/all/Gaussian/09-d01/g09/g09] [--gaussian_basis B3LYP
              [--metal_cutoff 2.8] [--metal_charges {MN:2, ZN:2, CA:2}]

Run or continue MD simulation. Allowed systems: Protein, Protein-Ligand, Protein-Cofactors(mult
```

How to run (minimal examples)

To run simulation:

run_md -p protein.pdb -l ligands.sdf --md_time 1

To extend simulation:

run_md --wdir_to_continue md_files/md_run/protein_ligand_*/ --md_time 2

G(P)BSA calculations:

run_gbsa --wdir_to_run md_files/md_run/protein_ligand_*

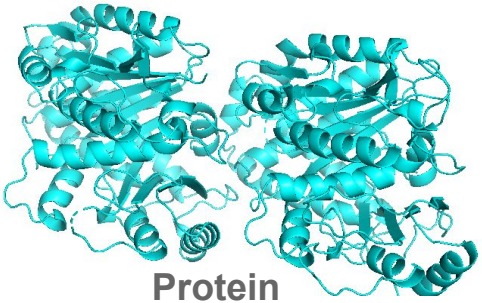
ProLIF calculations:

run_prolif --wdir_to_run md_files/md_run/protein_ligand_*

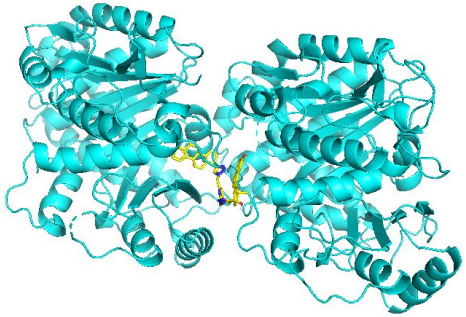
INSTITUTE OF MOLECULAR AND TRANSLATIONAL MEDICINE

Main features of StreaMD:

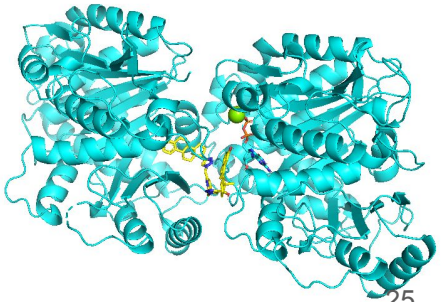
- **GROMACS engine**
- **Default set of optimal parameters for calculations, which can be customized**
- **Support of simulations of different molecular systems in explicit water solvent:**
  - protein only, protein-ligand, protein-cofactor(s), protein-ligand-cofactor(s)
- **Support of modeling of boron-containing molecules**
  - RESP charges calculation using Gaussian tool
- **Support of Building Bonded Model for A Ligand Binding Metalloprotein with MCPB.py**
  - to simulate proteins with specific metal ions not parametrized in commonly used FF
- **The ability to continue interrupted or to expand already finished simulations**
- **Support of distributed computing using Dask library across a network of servers or cluster**
- **GPU calculations support**



Protein



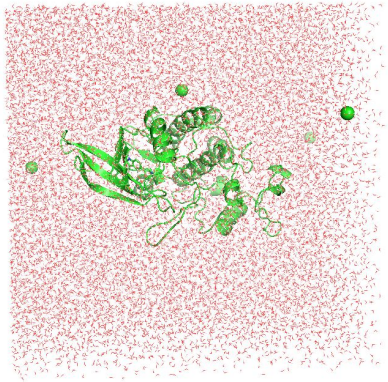
Protein-Ligand



Protein-Ligand-Cofactor

Main features of StreaMD:

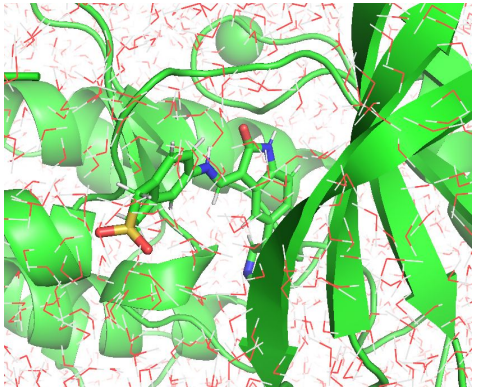
- **Automatic analysis of simulation:**
  - separate RMSD plots for protein, ligand and cofactors objects
  - a plot of flexibility of side chains of amino acids (RMSF)
  - a plot and a pdb file with radius of gyration
  - a single frame pdb file for the topology and a short subset of the trajectory for the quick visual inspection
  - a fitted trajectory (with removed periodic boundary conditions, aligned and centered on the first frame) to use for energy or protein-ligand interaction calculations
  - interactive trajectory convergence analysis for multiple complexes
- **Support of analysis by additional instruments:**
  - ***ProLIF***: Ligand-Protein interactions (time-dependent function, stability analysis)
    - *https://github.com/chemosim-lab/ProLIF*
  - ***gmx_MMPBSA:*** Calculation of Binding Energy by MM(PB)GBSA
    - *https://github.com/Valdes-Tresanco-MS/gmx_MMPBSA*



Check your own MD simulations

mdrun/md_files/md_run/protein_H_HIS_ligand

mdrun/md_files/md_run/protein_H_HIS_ligand/md_analysis



StreaMD Analysis Output Files

System stability & issues

potential.png

density.png

pressure.png

temperature.png

Minimization & Equilibration stages

Pose stability & System issues

rmsd_complex-name.png

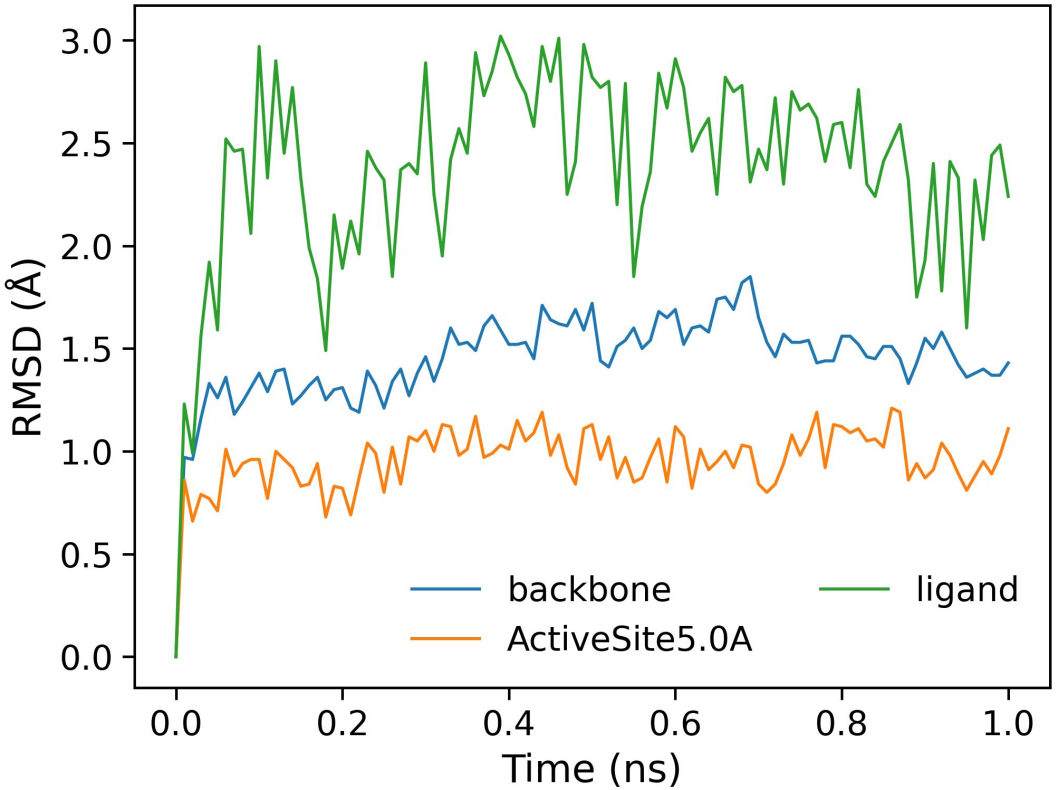
gyrate_complex-name.png

rmsf_complex-name.png

rmsd_mean_std_time-ranges_time.html

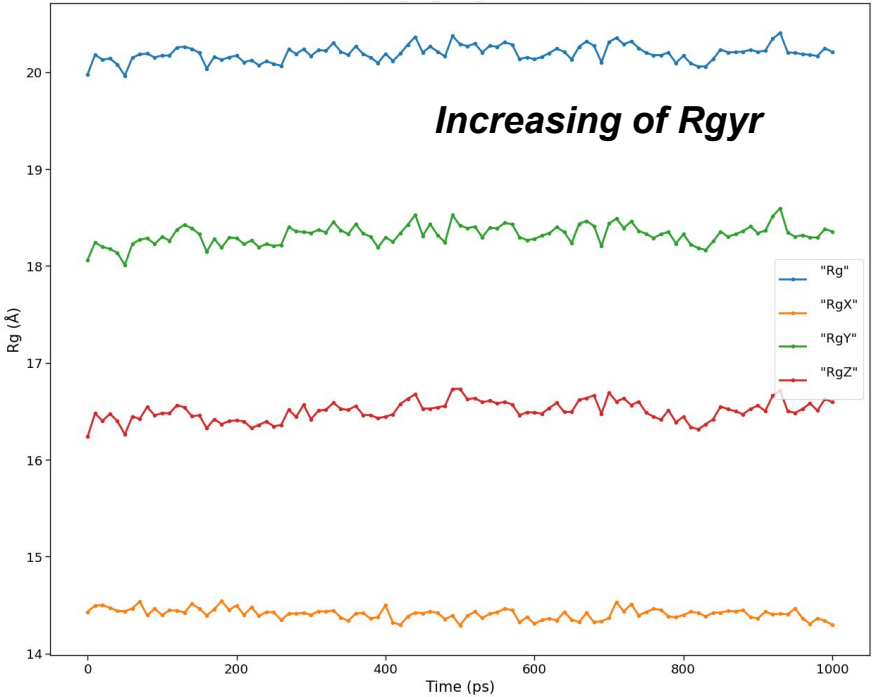
Production Simulation

Root mean square deviation of atomic positions

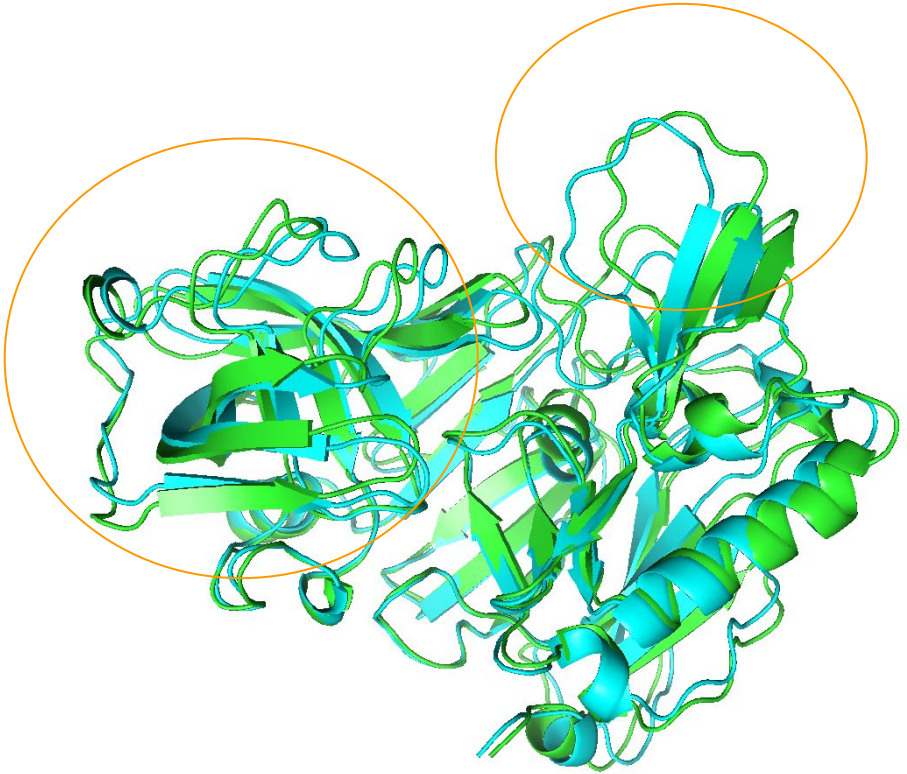


rmsd_*complex-name.png*

Radius of gyration (total and around axes)



gyrate_*complex-name.png*



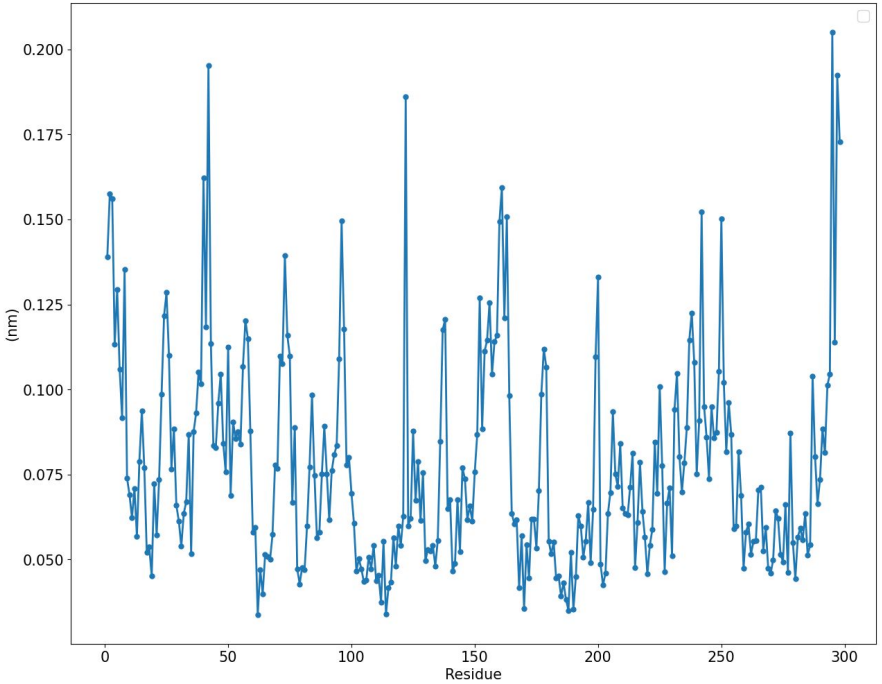
Conformer at the 1st ps
Conformer at the 1000th ps

Value of Rgyr:

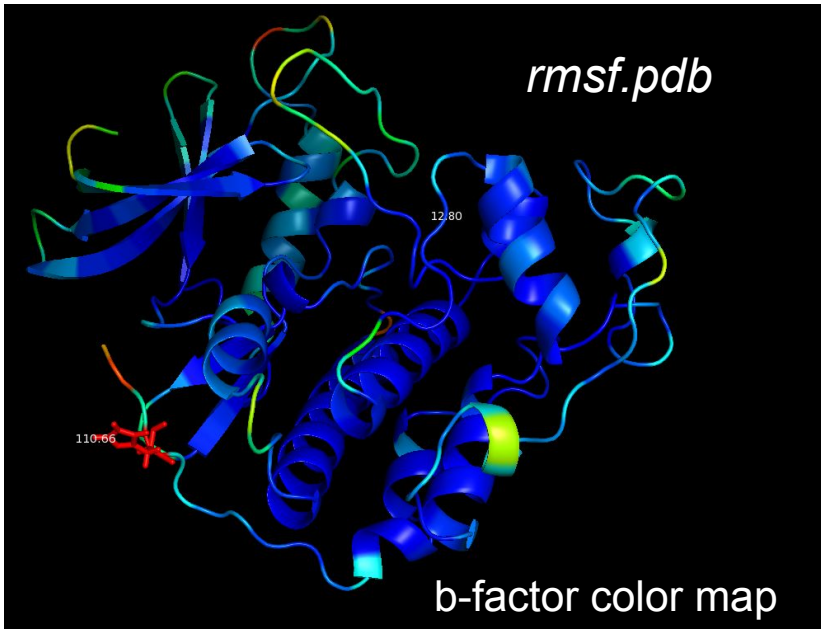
Decreasing - compression

Increasing - extension

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



rmsf_*complex-name.png*

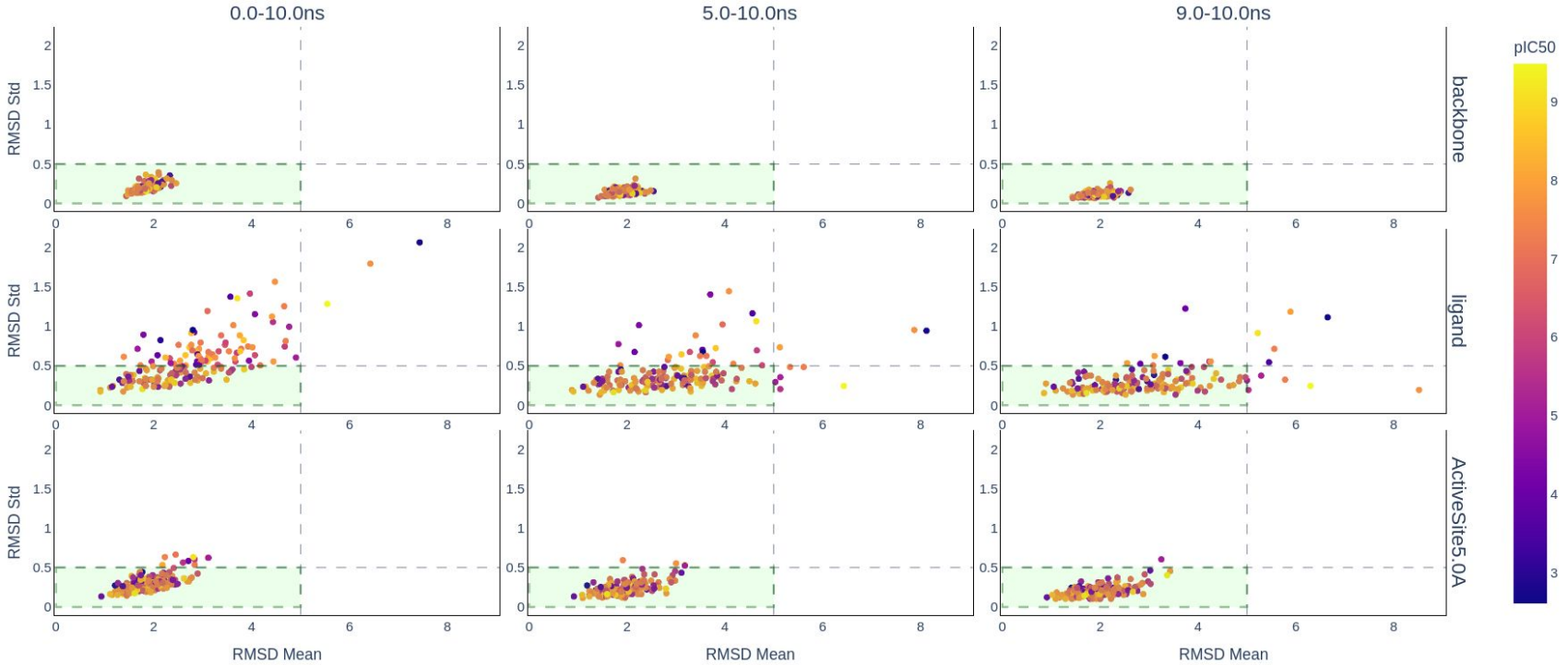


High RMSF: Loops or flexible regions. Disordered structure

INSTITUTE OF MOLECULAR AND TRANSLATIONAL MEDICINE

Trajectory convergence analysis

Interactive html files help to identify converged segments of trajectories



rmsd_mean_std_time-ranges_*time.html*

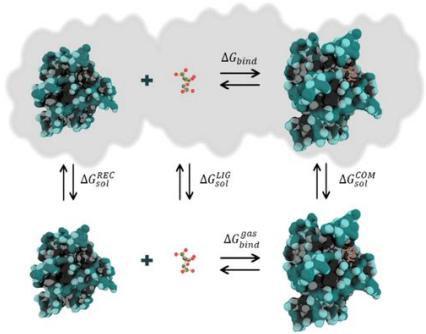
*The **average RMSD** provides insight into ligand movement or rotation relative to its initial pose, while the **standard deviation** reflects the stability of the ligand pose within the selected trajectory segment.*

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

MMPBSA / MMGBSA

- to estimate binding affinity of protein-ligand complexes

__Free binding energy__ is calculated as the difference in free energy between __the bound__ and __unbound__ states of the molecules. Where __the free energy__ is derived from:

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}
- **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/

```
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

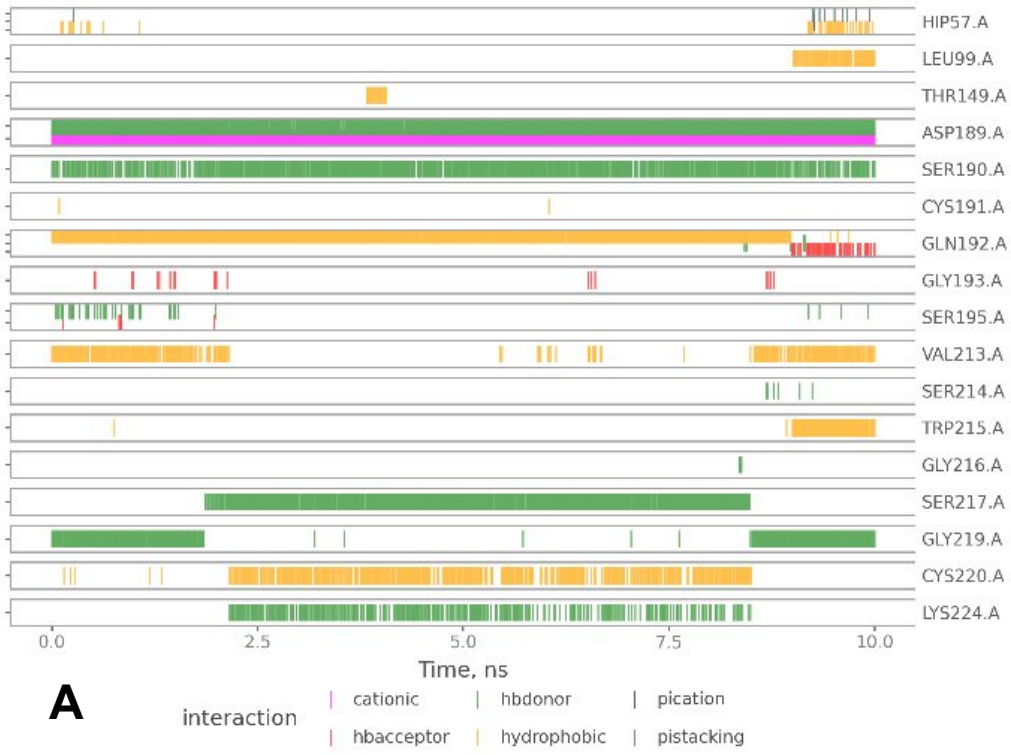
MMPBSA Energy and **MMGBSA Energy** cannot be compared within the different methods. But you can rank your molecules by energies obtained from each method separately.

ProLIF (Protein-Ligand Interaction Fingerprints)

run_prolif -i mdrun/md_files/md_run/protein_HIS_ligand/

INSTITUTE OF MOLECULAR AND TRANSLATIONAL MEDICINE

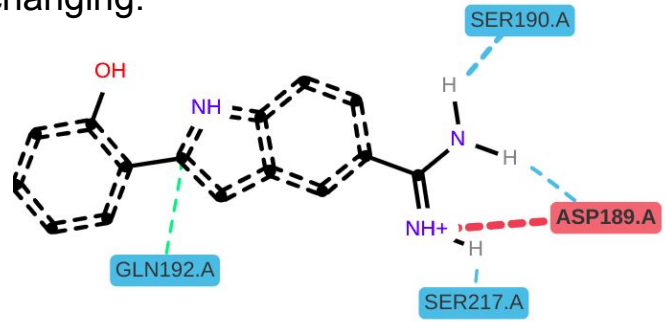
Protein-Ligand Interaction Fingerprints



Protein-ligand interactions detected for the trypsin dataset. (a) Occurrence of contacts in course of simulation detected for 1GI6 protein-ligand complex during 10 ns MD simulation.

What information can we retrieve from analysis of protein-ligand interactions?

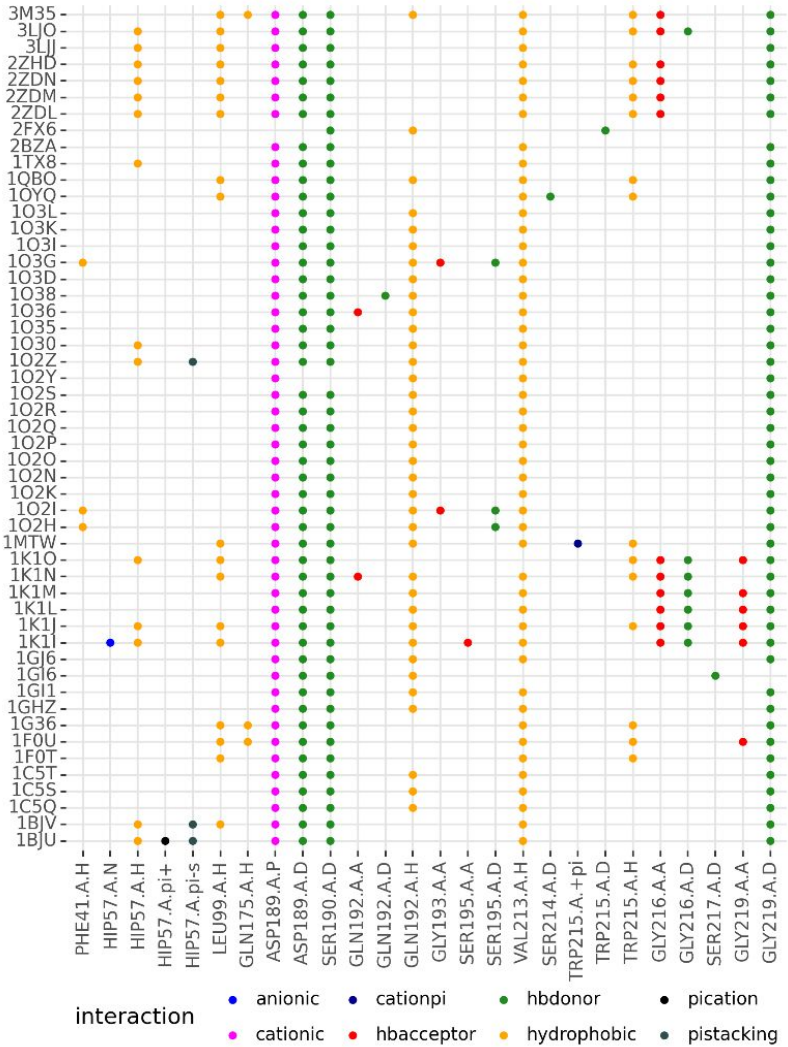
- By default, the tool creates 2 types of protein-ligand interactions analysis outputs
- The analysis of individual protein-ligand systems may show **which contacts are co-occurred** and **how these groups of contacts change during the simulation** that may suggest ligand moving or pose changing.



most frequently occurred ligand binding mode (>= 60% occupancy)

INSTITUTE OF MOLECULAR AND TRANSLATIONAL MEDICINE

Interaction fingerprints for the whole trypsin dataset occurred in at least 60% of frames of 10 ns MD trajectories

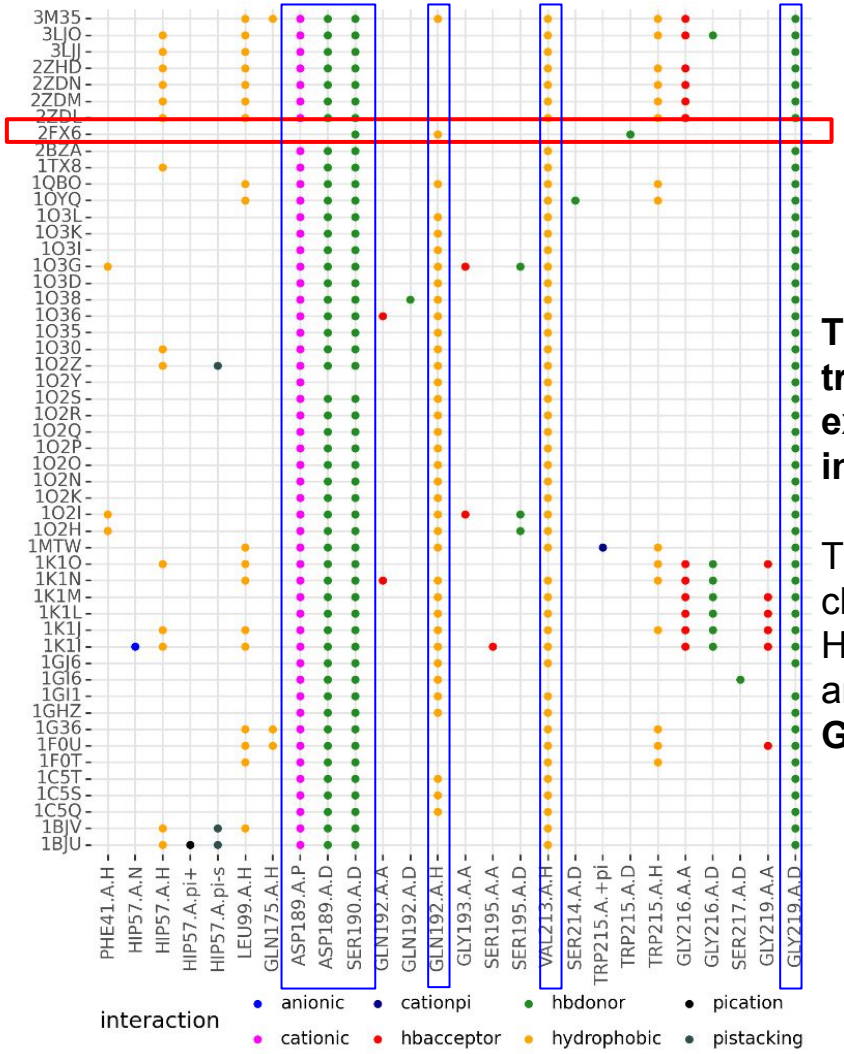


The analysis of contacts observed for multiple ligands may help **to identify the most frequently observed contacts (interaction patterns)** and **identify ligands which do not follow them**, that may indicate their unique binding modes or issues in a simulation setup.

A	B	C	D
Name	Frame	phe41.a.hydrophobic	hip57.a.hydrophobic
2BZA	0	FALSE	FALSE
1O2Y	0	FALSE	FALSE
1K1O	0	FALSE	TRUE
2ZDL	0	FALSE	FALSE
1QBO	0	FALSE	FALSE
1O2R	0	FALSE	FALSE
2ZHD	0	FALSE	TRUE
1G36	0	FALSE	FALSE

INSTITUTE OF MOLECULAR AND TRANSLATIONAL MEDICINE

*Ligand from **2FX6** complex did not follow this pattern.*



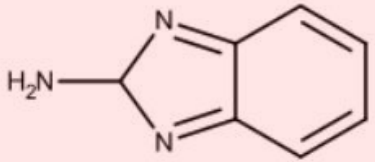
The analysis of the whole set of trypsin inhibitors revealed as expected the common interaction pattern.

The majority of ligands have charged interaction with **Asp189**, H-bonds with **Ser190** and **Gly219** and hydrophobic interactions with **Gln192** and **Val213**.

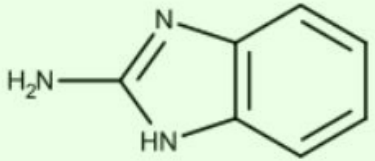
INSTITUTE OF MOLECULAR AND TRANSLATIONAL MEDICINE

Protein-Ligand Interaction Fingerprints

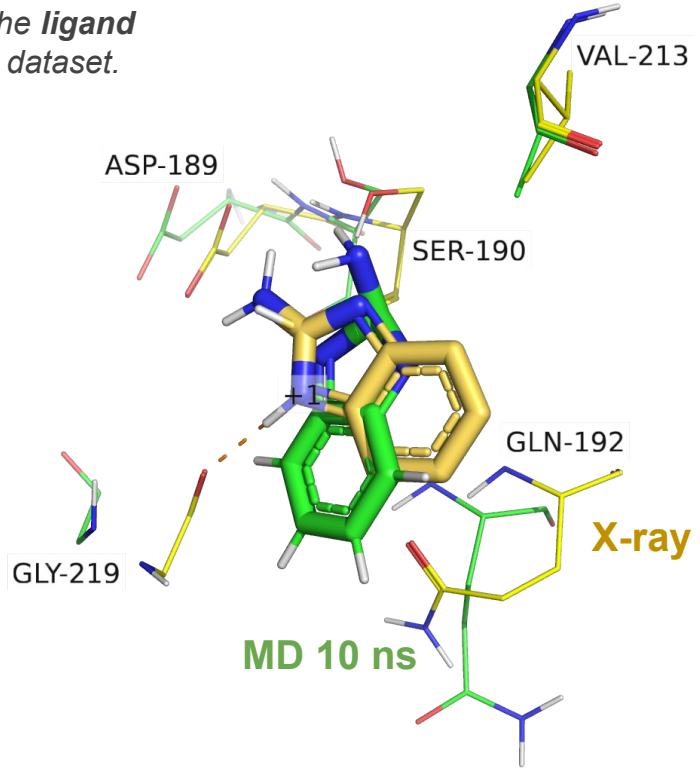
*Visual inspection of a ligand MD trajectory revealed that the structure of the **ligand was wrongly annotated in the PDB database** and was not fixed in the dataset.*



The structure from PDB database and data set of Bahia et al[23]



The structure reported in the original manuscript[26].



*The bond orders were incorrectly interpreted, that results in wrong geometry of the structure and that the **ligand started to move away from its initial pose** and could not form expected contacts.*

Availability and requirements

Project name: StreaMD

GitHub: https://github.com/ci-lab-cz/streamd

Operating system(s): Linux

Programming language: Python 3

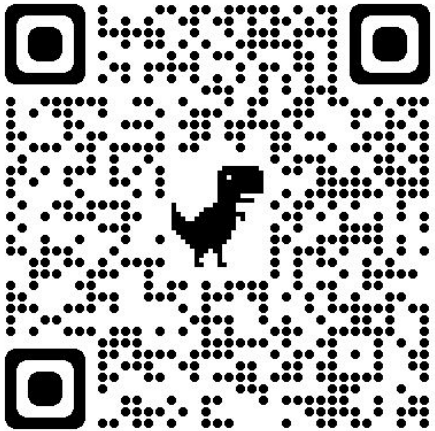
Other requirements: GROMACS, RDKit, ProLIF, Antechamber, MDAnalysis, Dask, Gaussian (optional, a license is required)

License: MIT

Any restrictions to use by non-academics: no

INSTITUTE OF MOLECULAR AND TRANSLATIONAL MEDICINE

Thank you for your attention!



https://github.com/ci-lab-cz/streamd

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic through INTER-EXCELLENCE II LUAUS23262, the e-INFRA CZ (ID:90254), ELIXIR-CZ (LM2018131, LM2023055), CZ-OPENSCREEN (LM2018130, LM2023052) grants and by European and Regional Fund project ENOCH (No. CZ.02.1.01/0.0/0.0/16_019/0000868).