Tutorial: Protein-Ligand Complex MD Setup with Jupyter Notebooks and BioBB

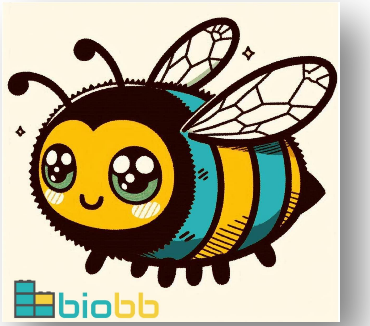
Exploring Biomolecular Modeling & Simulations 9-10/04/2025

EuroCC4 & BioExcel training workshop

Adam Hospital

adam.hospital@irbbarcelona.org

Agenda



Thursday, 10 April

10:00 → 11:00	**Interoperable and reproducible biomolecular simulation workflows using BioExcel Building Blocks (BioBB)** **Speaker:** Dr Adam Hospital (Institute for Research in Biomedicine in Barcelona (IRB-Barcelona) and Spanish National Institute of Bioinformatics (INB, ELIXIR-ES))	1h
11:00 → 11:30	Coffee Break	30m
11:30 → 13:00	**Hands on - Interoperable and reproducible biomolecular simulation workflows using BioExcel Building Blocks (BioBB)** **Speaker:** Dr Adam Hospital (Institute for Research in Biomedicine in Barcelona (IRB-Barcelona) and Spanish National Institute of Bioinformatics (INB, ELIXIR-ES))	1h 30m
13:00 → 14:00	Lunch Break	1h
14:00 → 15:00	**Fundamentals and practical use cases of free energy calculations with PMX** **Speaker:** Dr Sudarshan Behera (Max Planck Institute for Multidisciplinary Sciences, Goettingen)	1h
15:00 → 15:30	Coffee Break	30m
15:30 → 17:00	**Hands on - Fundamentals and practical use cases of free energy calculations with PMX** **Speaker:** Dr Sudarshan Behera (Max Planck Institute for Multidisciplinary Sciences, Goettingen)	1h 30m

Session 1 (lecture):

Interoperable and reproducible biomolecular simulation workflows using BioBB

- Workflows & Biomolecular workflows
- BioBB library
- BioBB workflows
  - Demonstration workflows (JN)
  - Pre-exascale workflows (HPC)

Session 2 (hands-on):

TRUBA
Türk Ulusal Bilim e-Altyapısı

Hands-on session on BioBB workflows:

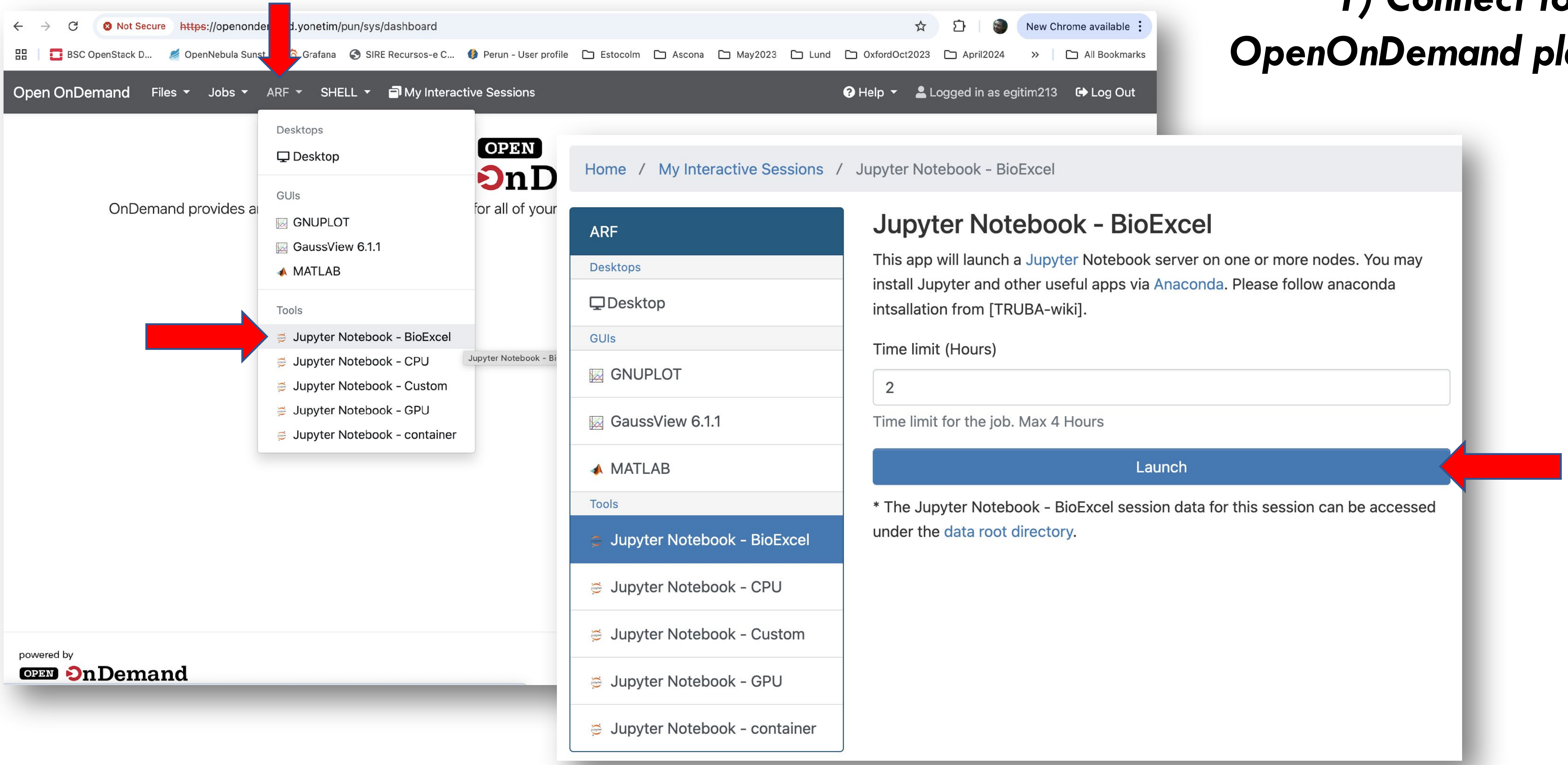
- BioBB demonstration workflow tutorial: GROMACS Protein-ligand complex MD setup
- Quick view on BioBB workflow collection
  - Jupyter Notebooks
  - Pure Python (High Throughput)

bioexcel

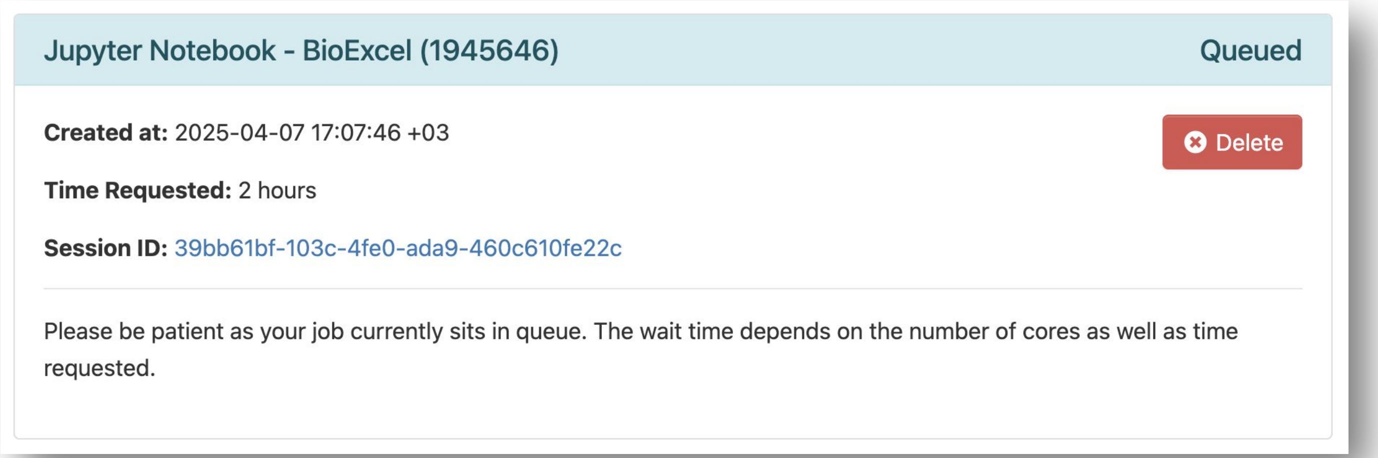
TRUBA OpenOnDemand

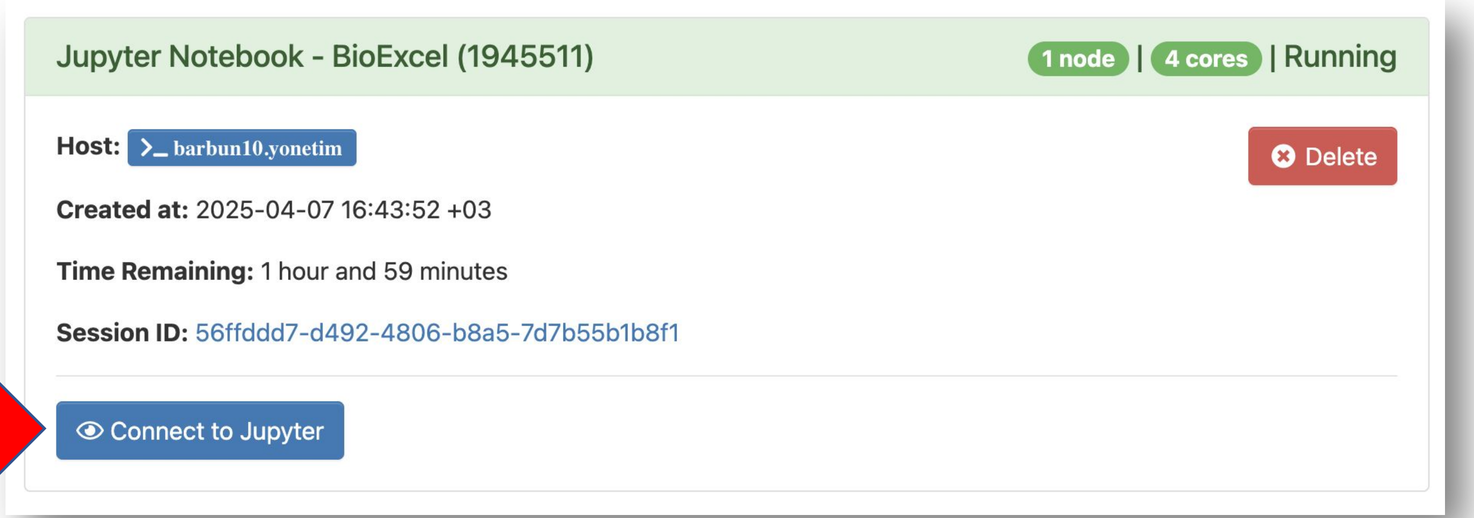
https://openondemand.yonetim/pun/sys/dashboard

1) Connect to OpenOnDemand platform



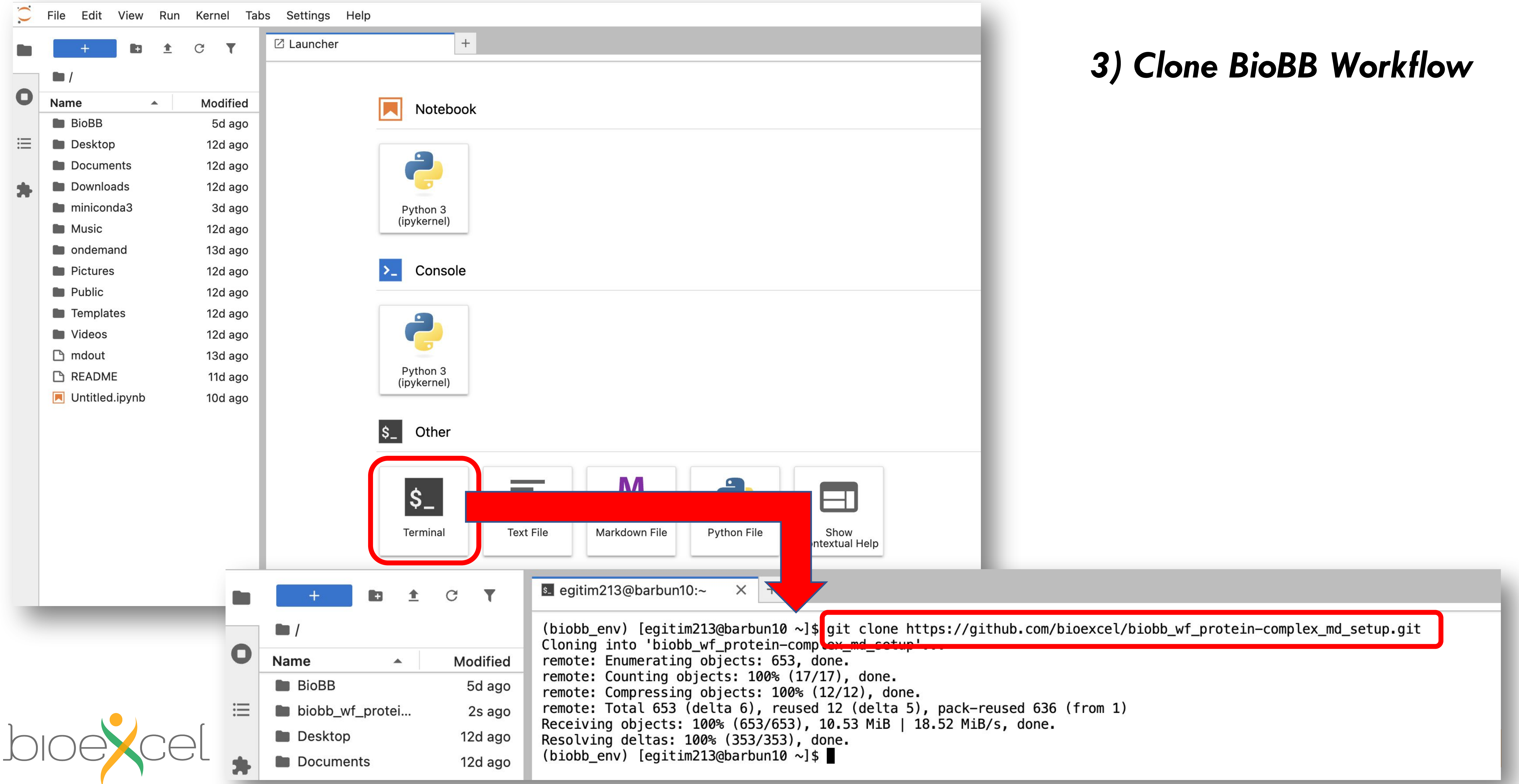
2) Jupyter Notebook - BioExcel



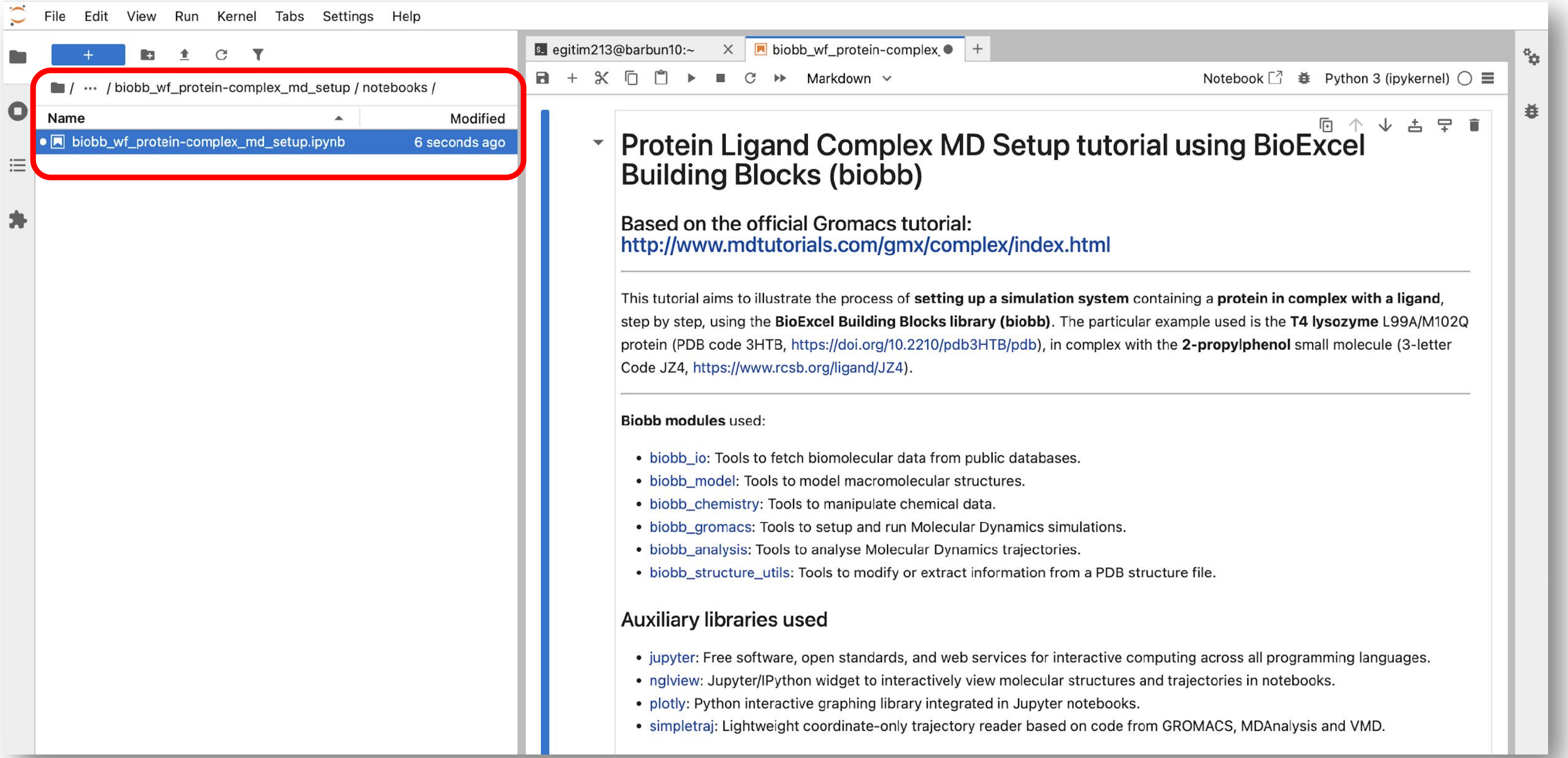


git clone https://github.com/bioexcel/biobb_wf_protein-complex_md_setup.git

3) Clone BioBB Workflow



4) Deploy BioBB Workflow Jupyter Notebook



Path: biobb_wf_protein-complex_md_setup/biobb_wf_protein-complex_md_setup/notebooks/biobb_wf_protein-complex_md_setup.ipynb

BioBB Demonstration Workflows

GROMACS PROTEIN-LIGAND COMPLEX MD SETUP 2024.1

This tutorial aims to illustrate the process of setting up a simulation system containing a protein in complex with a ligand, step by step, using the BioExcel Building Blocks library (biobb). The particular example used is the T4 lysozyme L99A/M102Q protein (PDB code 3HTB), in complex with the 2-propylphenol small molecule (3-letter Code JZ4).

WorkflowHub | Launch | Download

View tutorial | Open Github repository | Open documentation

gmx ligand md protein

(*) **MyBinder** provides a **free**, online version of **Jupyter Lab**. Take into account that the provided **resources** are **finite** and, in some occasions, it can take a long time to load or to execute your notebooks. **Please be patient** and don't try to execute several notebooks at the same time.

https://mmb.irbbarcelona.org/biobb/workflows

- *MD setup (Protein / DNA) (AMBER / GROMACS)*
- *Ligand parameterization*
- *Protein-Ligand Docking*
- *Free energy calculations*
- *DNA helical parameters*
- *Conformational Ensemble generation*

bioexcel

GROMACS Protein-Ligand MD Setup

GROMACS PROTEIN-LIGAND COMPLEX MD SETUP

2024.1

This tutorial aims to illustrate the process of setting up a simulation system containing a protein in complex with a ligand, step by step, using the BioExcel Building Blocks library (biobb). The particular example used is the T4 lysozyme L99A/M102Q protein (PDB code 3HTB), in complex with the 2-propylphenol small molecule (3-letter Code JZ4).

WorkflowHub | Launch | Download

View tutorial | Open Github repository | Open documentation

gmx ligand md protein

(*) MyBinder provides a **free**, online version of **Jupyter Lab**. Take into account that the provided **resources** are **finite** and, in some occasions, it can take a long time to load or to execute your notebooks. **Please be patient** and don't try to execute several notebooks at the same time.

- **Complex: T4 Lysozyme (3HTB) with 2-propylphenol (JZ4)**
- **Ligand parameterization (ACPype)**
- **GROMACS MD Setup (Min + NVT eq + NPT eq + unrestrained short MD)**

Questions (1)

YES =

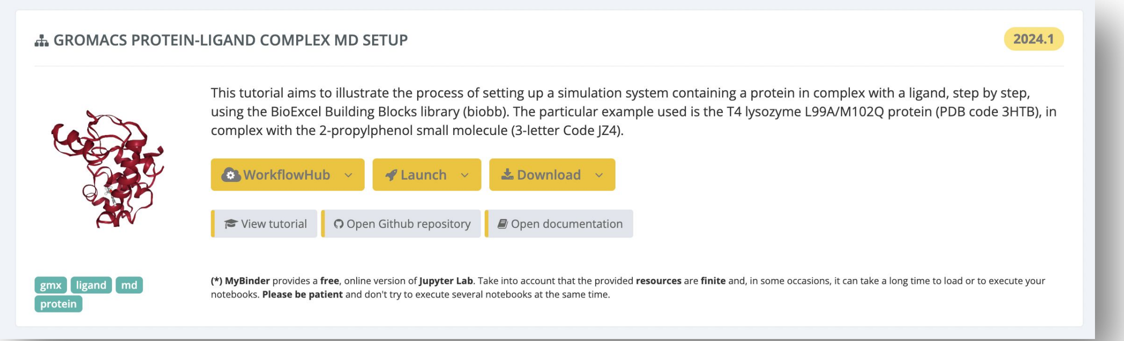
- ❑ **Have you ever setup/run a MD simulation using GROMACS before? (Y/N)**

- ❑ **Have you ever setup/run a MD simulation using other MD packages (AMBER, NAMD, DESMOND…)? (Y/N)**

- ❑ **Have you ever setup/run a MD simulation of a protein-ligand complex? (Y/N)**

Index

- **Part 1: Install and launch the workflow**
  - Conda Environment
  - Jupyter Notebook
- **Part 2: Download PDB and fixing structure**
- **Part 3: Generating topologies**
  - Protein
  - Ligand
- **Part 4: Generating protein-ligand complex structure**
  - Ligand Position Restraints
  - MD ready structure
- **Part 5: System Topology**
- **Part 6: MD Setup & Run**
- **Part 7: Trajectory (basic) analyses**



Questions (2)

YES =

- ❑ **Are you familiar with Linux OS? (Y/N)**
- ❑ **Have you ever used Conda packages/environments before? (Y/N)**
- ❑ **Have you ever used Jupyter Notebooks before? (Y/N)**

Part 1: ~~Install and~~ launch the workflow

CONDA® READY

Using TRUBA resources:

1) Connect to **OpenOnDemand** platform
2) Open **Jupyter Notebook** – BioExcel
3) Clone **BioBB Workflow**
4) Deploy **Jupyter Notebook**

At home:

CONDA® REQUIRED

Conda Installation and Launch

```
git clone https://github.com/bioexcel/biobb_wf_protein-complex_md_setup.git
cd biobb_wf_protein-complex_md_setup
conda env create -f conda_env/environment.yml
conda activate biobb_wf_protein-complex_md_setup
jupyter-notebook biobb_wf_protein-complex_md_setup/notebooks/biobb_wf_protein-complex_md_setup.ipynb
```

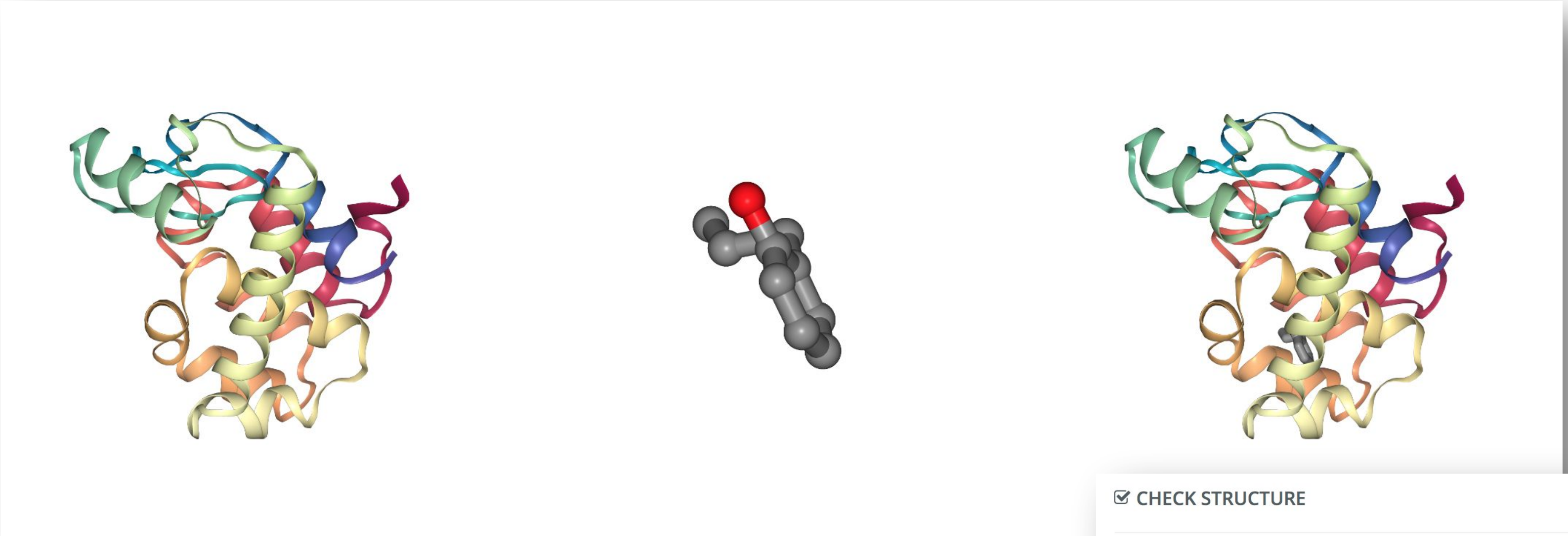
AnaConda3 installation: https://mmb.irbbarcelona.org/biobb/availability/tutorials

Questions (3)

YES =

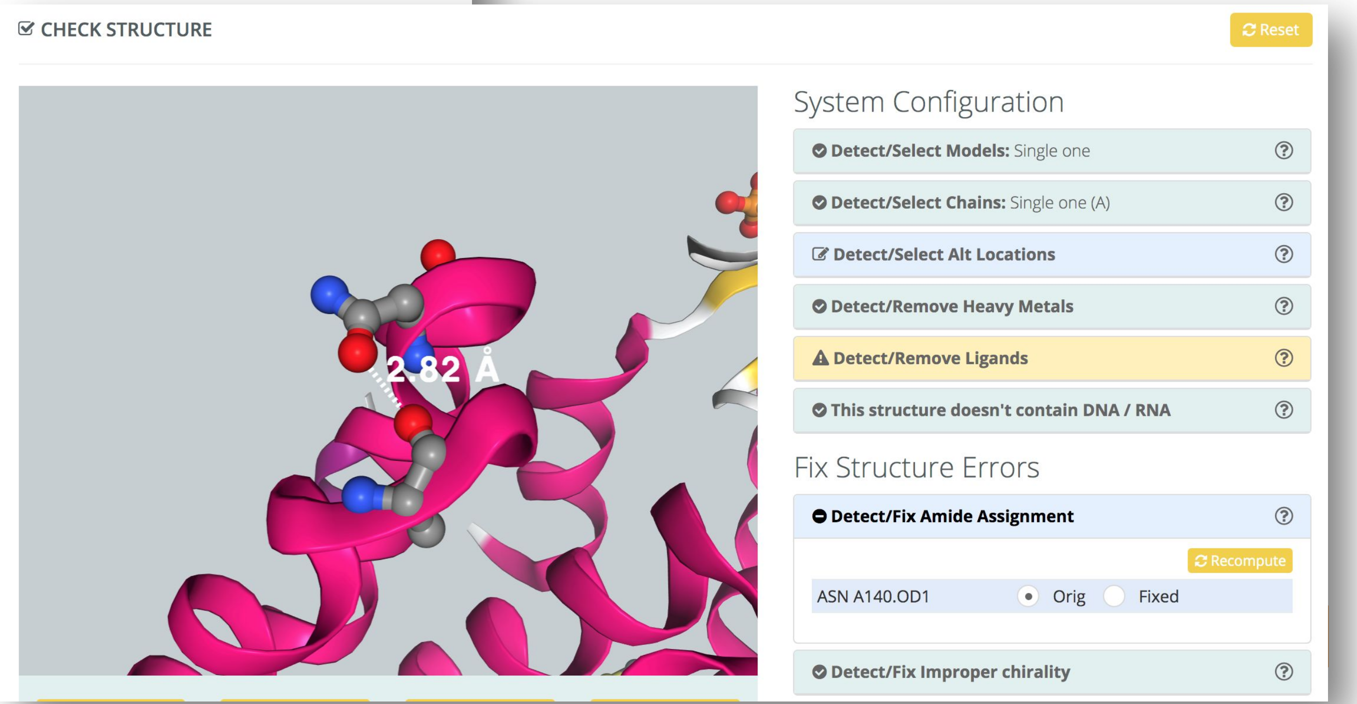
- Have you been following so far? (Y/N)
- Are you familiar with PDB files and their content? (Y/N)
- Do you use graphical interfaces in your everyday work? (Y/N)

Part 2: Download PDB, check the structure, and generate protein topology



1. Input Parameters
2. Fetching PDB Structure
3. Fix Protein Structure
4. Create Protein System Topology

https://mmb.irbbarcelona.org/biobb-wfs/structure/step1



Questions (4)

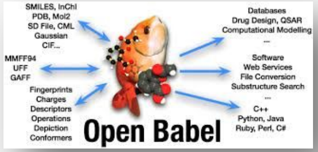
YES =

- ❑ **Have you been following so far? (Y/N)**
- ❑ **Have you ever parameterized a small molecule? (Y/N)**

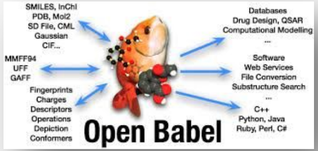
Part 3: Parameterizing small molecule (topology)

[2]:
```python
# Ligand: Download ligand structure from MMB PDB mirror REST API (https://mmb.irbbarcelona.org/api/)
# Import module
from biobb_io.api.ligand import ligand

# Create prop dict and inputs/outputs
input_structure = ligandCode + '.pdb'

prop = {
    'ligand_code' : ligandCode
}

#Create and launch bb
ligand(output_pdb_path=input_structure,
       properties=prop)
```

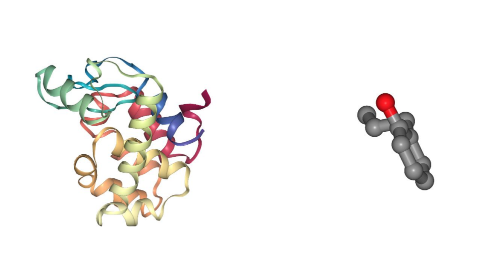
[4]:
```python
# Babel_add_hydrogens: add Hydrogen atoms to a small molecule
# Import module
from biobb_chemistry.babelm.babel_add_hydrogens import babel_add_hydrogens

# Create prop dict and inputs/outputs
output_babel_h = ligandCode + '.H.mol2'

prop = {
    'ph' : pH,
    'input_format' : 'pdb',
    'output_format' : 'mol2'
}

#Create and launch bb
babel_add_hydrogens(input_path=input_structure,
                    output_path=output_babel_h,
                    properties=prop)
```



[6]:
```python
# Babel_minimize: Structure energy minimization of a small molecule after being modified adding hydrogen atom
# Import module
from biobb_chemistry.babelm.babel_minimize import babel_minimize

# Create prop dict and inputs/outputs
output_babel_min = ligandCode + '.H.min.pdb'
prop = {
    'method' : 'sd',
    'criteria' : '1e-10',
    'force_field' : 'GAFF'
}

#Create and launch bb
babel_minimize(input_path=output_babel_h,
               output_path=output_babel_min,
               properties=prop)
```



[9]:
```python
# Acpype_params_gmx: Generation of topologies for GROMACS with ACPype
# Import module
from biobb_chemistry.acpype.acpype_params_gmx import acpype_params_gmx

# Create prop dict and inputs/outputs
output_acpype_gro = ligandCode + 'params.gro'
output_acpype_itp = ligandCode + 'params.itp'
output_acpype_top = ligandCode + 'params.top'
output_acpype = ligandCode + 'params'
prop = {
    'basename' : output_acpype,
    'charge' : mol_charge
}

#Create and launch bb
acpype_params_gmx(input_path=output_babel_min,
                  output_path_gro=output_acpype_gro,
                  output_path_itp=output_acpype_itp,
                  output_path_top=output_acpype_top,
                  properties=prop)
```



5. Create ligand system topology

Questions (5)

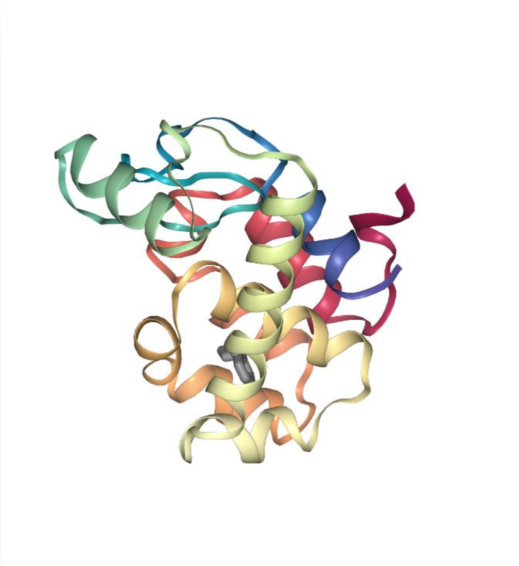
YES =

- ❑ **Have you been following so far? (Y/N)**
- ❑ **Are you familiar with the concept of restraints in MD? (Y/N)**
- ❑ **Are you familiar with the concept of force field parameters? (Y/N)**

Part 4: Generating protein-ligand complex structure

6. Preparing Ligand Restraints

7. Create new protein-ligand complex structure file



Ligand Restraints

Terms	Abbreviation
Molecule and residue information	units
Atom names	names
Atom types	types
Atomic charges	chargs
Atomic connectivities	connects
Atomic coordinates	coords
Atomic masses	masses
Bonded parameters (bond, angle, dihedral)	bond params
Nonbonded parameters (electrostatic, VDW)	nonb params

Structure – Topology atom names matching

bioexcel

Questions (6)

YES =

- ❑ **Have you been following so far? (Y/N)**

- ❑ **Have you ever setup/run a MD simulation of a protein-ligand complex using GROMACS? (Y/N)**

Part 5: Generating protein-ligand complex topology

8. Create new protein-ligand complex topology file

```
; Include forcefield parameters
#include "amber99sb-ildn.ff/forcefield.itp"

; Including ligand ITP
#include "JZ4params.itp"

; Ligand position restraints
#ifdef POSRES_JZ4
#include "JZ4_posres.itp"
#endif


[ moleculetype ]
; Name            nrexcl
Protein_chain_A     3
```

```
; Include topology for ions
#include "amber99sb-ildn.ff/ions.itp"

[ system ]
; Name
Protein

[ molecules ]
; Compound        #mols
Protein_chain_A     1
JZ4params           1
```

Questions (7)

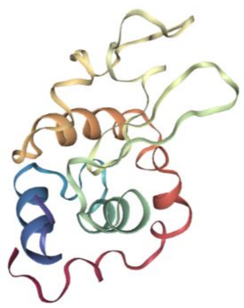
YES =

- ❑ **Have you been following so far? (Y/N)**
- ❑ **Did you follow the previous GROMACS hands-on session? (Y/N)**

Part 6: MD Setup

GROMACS PROTEIN MD SETUP

2024.1



This tutorial aims to illustrate the process of setting up a simulation system containing a protein, step by step, using the BioExcel Building Blocks library (biobb). The particular example used is the Lysozyme protein (PDB code 1AKI).

WorkflowHub | Launch | Download

View tutorial | Open Github repository | Open documentation

gmx md protein

(*) MyBinder provides a **free**, online version of **Jupyter Lab**. Take into account that the provided **resources** are **finite** and, in some o
notebooks. **Please be patient** and don't try to execute several notebooks at the same time.

9. Create Solvent Box

10. Fill the Box with Water Molecules

11. Adding Ions

12. Energetically Minimize the System

13. Equilibrate the System (NVT)

14. Equilibrate the System (NPT)

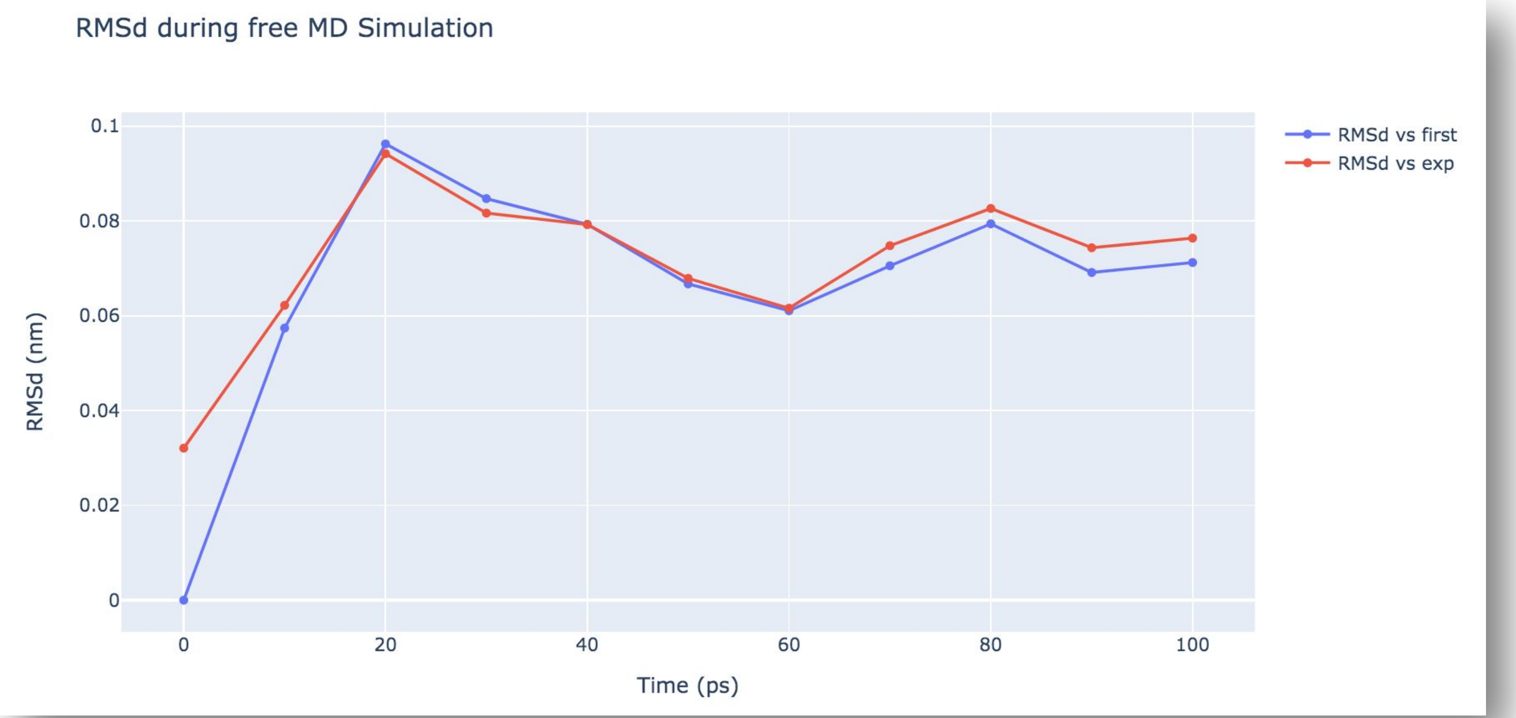
15. Free Molecular Dynamics Simulation

Questions (8)

YES = 

- ❑ **Have you been following so far? (Y/N)**
- ❑ **Are you familiar with GROMACS analysis tools? (Y/N)**
- ❑ **Are you familiar with imaging process (periodicity issues)? (Y/N)**

Part 7: Trajectory post-processing



Final Questions

YES =

- ❑ **Have you been able to follow the tutorial? (Y/N)**
- ❑ **Would you like to know more about the BioExcel Building Blocks? (Y/N)**

Follow up:

- *https://ask.bioexcel.eu/c/BioExcel-Building-Blocks-library*
- *https://github.com/bioexcel/biobb/issues*

Acknowledgments

INSTITUTE FOR RESEARCH IN BIOMEDICINE

IRB BARCELONA

Federica Battistini

Genís Bayarri

Modesto Orozco

thank you

Barcelona Supercomputing Center

Centro Nacional de Supercomputación

Pau Andrio

Josep Ll. Gelpí

bioexcel

Centre of Excellence for Computational Biomolecular Research

BioExcel Center of Excellence, funded from the European Union's Horizon 2020 Framework Programme for Research and Innovation under Specific Grant Agreements No. 675728, 823830, 101093290 (BioExcel-1, BioExcel-2 and BioExcel-3).

Co-funded by the European Union

EuroHPC Joint Undertaking