Molecular dynamics of proteins in the cloud

Antonio Rosato
Magnetic Resonance Center
University of Florence
rosato@cerm.unifi.it
CERM/CIRMMP
Magnetic Resonance Center

An INSTRUCT Core Center

Competence center “Ivano Bertini” (2015)

Biobank

Mass Spectrometry Conference Room

X-Ray Crystallography Library

He Liquifier Workstations

Relaxometer (0.01-40 MHz)

Molecular biology, cellular biology and biophysics labs

www.cerm.unifi.it

for metabolomics

CERM/CIRMMP
Magnetic Resonance Center
The initial steps of collagenolysis. (A) Closed (left) and open/extended (right) forms of FL-MMP-1 in equilibrium. (B) The extended protein binds THP chains 1T-2T at Val23-Leu26 with the HPX domain and the residues around the cleavage site with the CAT domain. The THP is still in a compact conformation. (C) Closed FL-MMP-1 interacting with the released 1T chain (in magenta). (D) After hydrolysis, both peptide fragments (C- and N-terminal) are initially bound to the active site. (E) The C-terminal region of the N-terminal peptide fragment is released.
Timescales of protein dynamics

- Global reorientation
- Water exchange
- Conformational exchange
- Motions of sidechains
- Folding/unfolding

- Fast librations

- Heteronuclear NOE, $T_1$ and $T_2$ relaxation
- CPMG and $T_1p$ experiments
- NMRD measurements
- RDC data

- Log(time/s)
Molecular dynamics simulations

Because no experimental technique can be applied to the investigation of all atoms/groups in a protein, computer simulations are needed to obtain a complete picture.

MD simulations are based on the integration of classical equations of motion (variation of position, velocity and acceleration) over small intervals of time $\Delta t$ (constant acceleration, linear motion). At each time step the system energy and forces over atoms are re-evaluated.

For protein MD simulations $\Delta t = 1-2$ fs. So for 1 $\mu$s of simulation, 1 billion steps are needed.

Calculations require several days and generate 10’s Gb.
MD simulations on GPGPUs

The ferritin homopolymer

TOTAL MW 480 kDa
24 subunits 175 aa each
Octahedral (432) symmetry

8 nm
12 nm

Ferritin plus solvent: 178910 atoms
Ion escape from a protein nanocage
MD vs. NMR-derived dynamics

IRED analysis of a 1 μs simulation of ubiquitin, compared to order parameters derived from experimental data

Pearson coefficient Exp-Calc = 0.82

Structure of our current portal for short MD calculations

Building a cloud-based web server for MD analysis

We used solutions developed by the INDIGO-DataCloud project (https://www.indigo-datacloud.eu/)

The AmberTools software package for MD was implemented as a container for Docker

We additionally implemented the Oneclient application in the Docker image

Onedata is a Dropbox-style cloud storage solution. Its Oneclient component allows mounting your Onedata space in the local file system tree
In the Dockerfile

```bash
[...] ENV PROVIDER_HOSTNAME=oneprovider.cloud.cnaf.infn.it
ENV ONECLIENT_AUTHORIZATION_TOKEN='AADDFF.......'
ENV MOUNT_POINT=/tmp/onedata
RUN mkdir /tmp/onedata

[...] EXPOSE 8080
CMD oneclient --authentication token $MOUNT_POINT --no-check-certificate && mvn jetty:run
```

Web server

Based on https://hub.docker.com/r/indigodatalcloudapps/ambertools-oneclient/
Building a cloud-based web server for MD analysis

The container further includes

• A data browser, so the container can be accessed via your web browser (i.e. no need to use the Onedata web interface in another tab).
• A viewer for macromolecular structures (NGL, a highly memory-efficient and scalable WebGL-based viewer)

We then instantiate the VM using Ansible with specifications in YAML

This is compatible with the PaaS Orchestrator of INDIGO
To instantiate the VM

YAML file

[...]  
- name: build the image  
  docker_image:  
    path: /tmp/site  
    name: cirmmp/mds2  

As we are in dev, the container is stored locally

- name: run the site in a docker container  
  docker_container:  
    name: mds2  
    privileged: True  
    image: "cirmmp/mds2"  
    dns_servers: 150.217.1.32  
    published_ports: 8080
How it looks

You can drag and drop files here, and they will go into your Onedata space. Implies passing the Onedata access token when the VM is set up.
Display of the results of MD analysis

Output data remain in your Onedata space when the VM is shut down
Infrastructure to compute MD simulations

Presently:
• GPGPU-based using the HTC infrastructure of EGI

Next
• Dockerized version for GPGPU VMs

Computing MD simulations is much more efficient on GPGPUs: see https://documents.egi.eu/public/ShowDocument?docid=2774
Instead, MD analysis does not require GPGPUs
Next steps

Integrate this with the INDIGO IAM (Identification and Authorization Mechanism), to simplify management of the access token

Adapt to check running calculations on the fly

Improve the design of the web interface
Thank you for your attention!

People involved:

• Davide Sala
• Andrea Giachetti

Financial support:
Italian Ministry of University
Regione Toscana
European Commission
R1 relaxation parameters of Menkes protein

MD simulations of the WT- and A629P-MNK6 domain in the apo-form.

After 0.5 µs of simulation the IRED analysis was performed

R1 correlations (Exp-Calc)

A629P = 0.50

WT = Sample partly aggregated