MD Simulations

Toni Giorgino
National Research Council of Italy
toni.giorgino@cnr.it
www.giorginolab.it

GitHub
@giorginolab

Thesis projects available

University of Padova for Prof. Fuxreiter
May 2024

https://github.com/giorginolab/MD-Tutorial-Data

This class

- Molecular dynamics is a powerful tool for studying molecular systems
- OpenMM is a software library that allows for efficient and customizable MD simulations
- It's exemplary of a modern well-maintained open-source library:
  - CI infrastructure, developed on GitHub
  - C++ w/ Python bindings
- We'll use the latter, testing *live* on Google Colab.

Molecular Dynamics

What is MD?

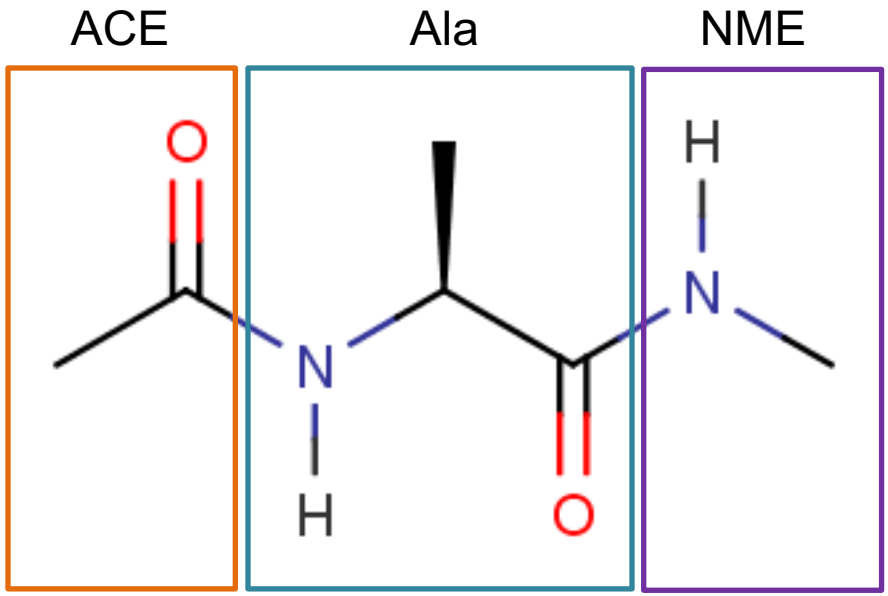
- Attempt the most detailed description of a system which is
  1. atomistic
  2. classical
- Model the internal *forces*…
- …in order to *integrate* the motion
- Hope in convergent *sampling*

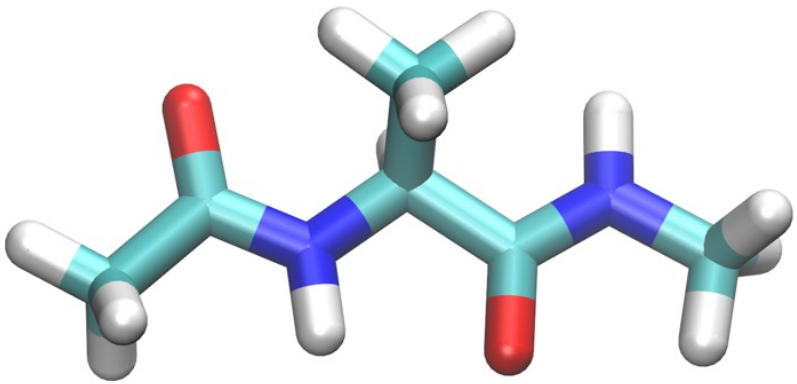
$$\vec{F}_i(\mathbf{x}) = m_i \ddot{\mathbf{x}}_i$$

Assumptions

- In this tutorial we shall deal with **unbiased** sampling approaches with **explicit** solvent, i.e.
  - no added forces except the "physical" ones in your system;
  - all of the system (including water molecules) have atomic resolution.
- Also, current classical MD does not address, by design, the following:
  - Chemical reactions, e.g. catalysis, phosphorylation, ubiquitination etc.
  - Protonation changes
- Finally, small molecules pose distinct challenges and need a separate, expensive **parameterization** step.

Alanine “dipeptide”

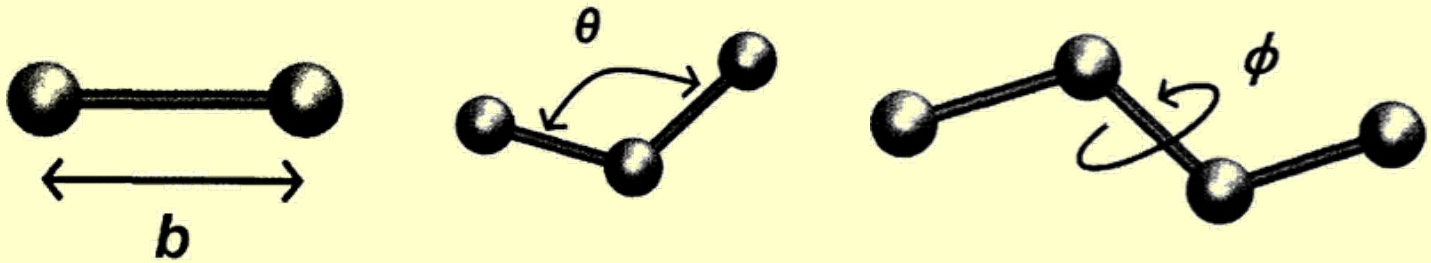




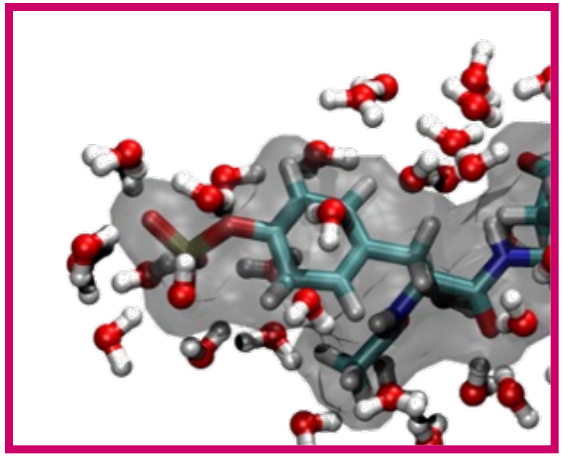
Bonded energy terms

+ Electrostatics

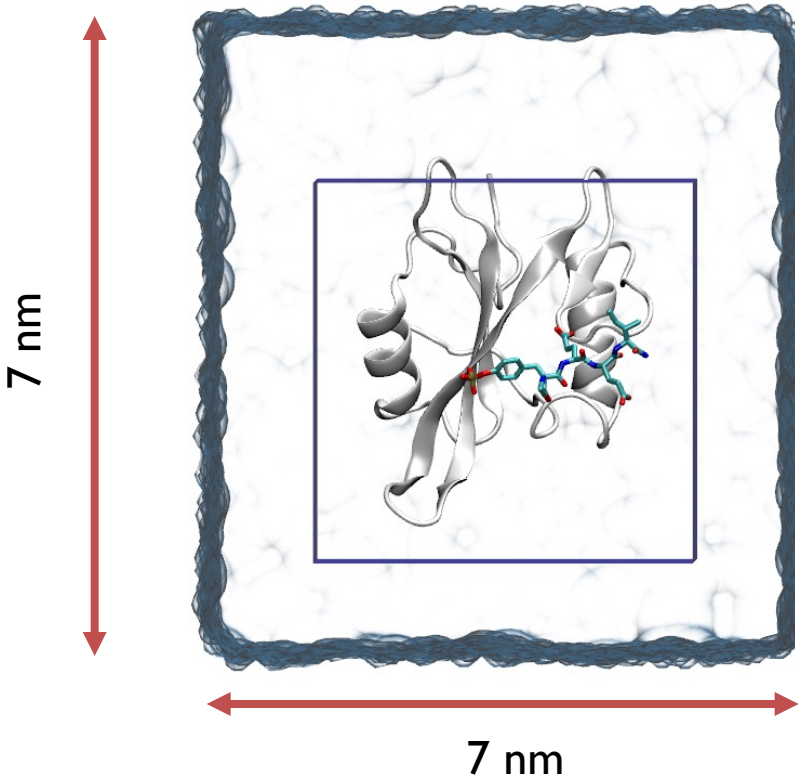
+ VdW

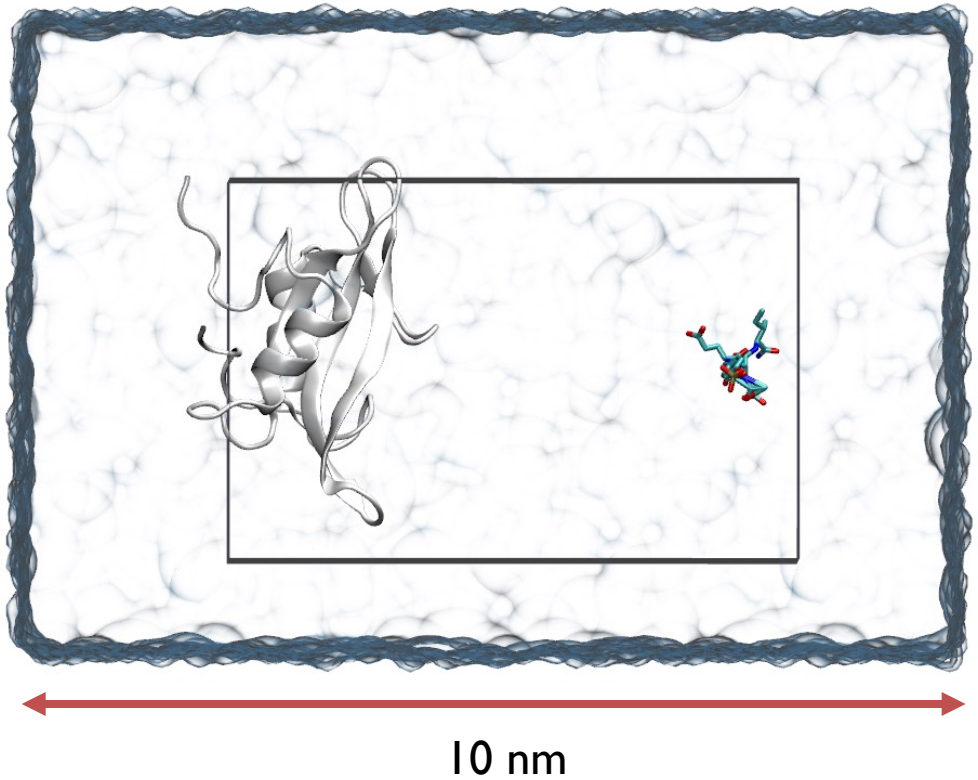


The forcefield is a database of interatomic parameters



- **Explicit solvation**
- → O(10^5) atoms
- **Unbiased dynamics**
- Update every 10^{-15} s (1 fs)





Event **≡ Binding / Unbinding / Folding / Unfolding / …**

* $1/t_{on}$ = association rate of SH2-pYEEI × [pYEEI]

Large gain

Ability to "play" biomolecular processes at *all-atom* resolution *in silico*

Molecular bases of folding, binding, selectivity, gating…

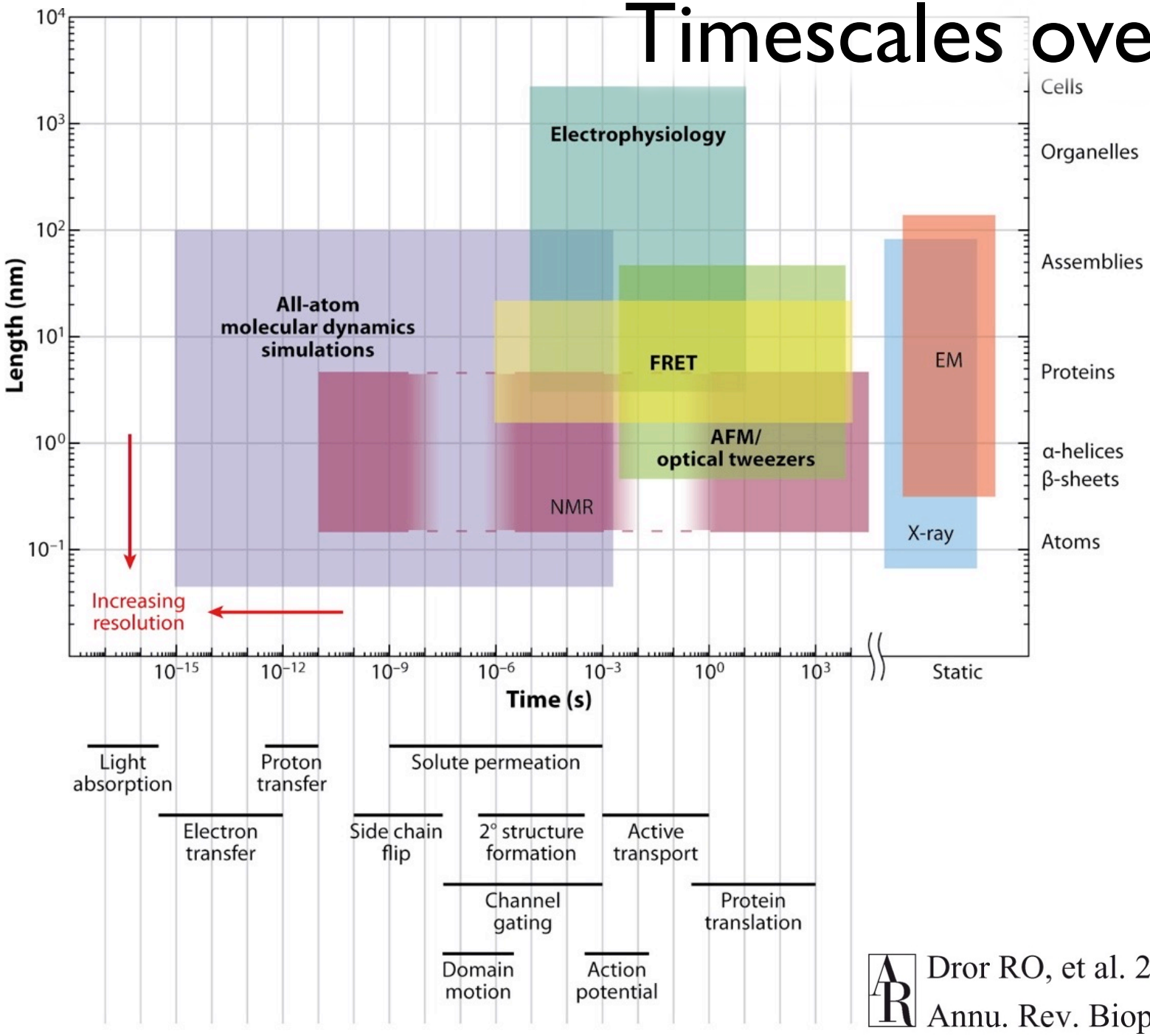
Large cost

E.g.*: t_{on} ~ 30 μs →
→ $\mathbf{10^{10}}$ integration timesteps →
→ 15 years single-CPU compute time

MD is entirely about timescales

- Your ability to obtain quantitative results is severely limited by the sampling ability you have. You will only be able to reach phenomena occurring on the sampled timescales, or shorter.
  - Sidechain rearrangements, diffusion-limited processes: usually possible *
  - Local flexibility: usually possible *
  - Membrane environments: ok-ish
  - Binding: hard but not impossible
  - Folding: very hard but not impossible
    - [*] Unless there are significant barriers.

Timescales overview



Dror RO, et al. 2012.
Annu. Rev. Biophys. 41:429–52

Patience and limits

- The following factors affect the running speed (usually expressed in ns per simulation day, ns/day)
  - System size. Reasonable is 100 AA ~ 30,000 atoms.
  - Computer speed. Forget laptops.
  - Definitely use GPUs.
  - Software.

General workflow

- Know your system
- Build
  - Cleanup
  - Assign forcefield
  - Solvate
  - Ionize
- Minimize
- Equilibrate
- Run

Forcefields

They run in “families”

- They are “databases” of atomic parameters
- We deal with non-polarizable all-atom ones: AMBER and CHARMM
  - There are many others, these are s.o. art
- Most notable difference: molecular types supported (lipids, drugs, …)
- They differ in (software) build procedures
- However OpenMM unifies them

The CHARMM family

- Originally coupled with the CHARMM software (MacKerell, Karplus), but independent
- Variants of note: C36M
  - toppar_c36_jul22.tgz
- Based on RTF (templates) and PRM (parameters) files
- Build: **psfgen** -> xxx.psf

CHARMM
Chemistry at HARvard Macromolecular Mechanics

- https://www.academiccharmm.org/
- https://mackerell.umaryland.edu/charmm_ff.shtml

The AMBER family

- Originally coupled with the Amber software (Case, Merz, Kollmann, …), but independent
- Variants of note: ff19SB
  - https://ambermd.org/AmberTools.php
- Based on **tleap** + its database
- Build: **tleap** --> xxx.prmtop

FF choice

Largely equivalent, i.e. subtle differences only appear at late stages.
FF choice is mostly due to the species in the system that one intends to model.

	CHARMM	AMBER
Proteins, peptides	++	++
Water, ions	++	++
Lipids (membranes)	++	+
Small molecules	+ (CGenFF)	++ (GAFF2)
Post-translational modif. *	+/-	+/-
Non-standard charge states	+/-	+/-
DNA	-	-
RNA	--	--
Non-standard AAs *	--	--

Somewhat subjective!

* Check individually

An example for Amber

Molecule/Ion Type	Force Field
protein	ff19SB
DNA	OL21
RNA	OL3
carbohydrates	GLYCAM_06j
lipids	lipids21
organic molecules (usually ligands)	gaff2
ions	•should be matched to water model; see force fields for ions for further discussion
water model	•should be matched to atomic ions; common water models include tip3p, spc/e, tip4pew, and OPC

System building

The build procedure

- It used to be somewhat convolved
- Generally
  - take PDB coordinates
  - filter out unwanted species
  - solvate
  - ionize
- Automation was (it still is) challenging
- OpenMM unified the build process (but it is still possible to use the old tools)

OpenMM

OpenMM

High performance, customizable molecular simulation.

.org

- OpenMM is a molecular dynamics simulation toolkit that allows for high-performance simulations of biomolecules.
- Allows for simulation of a variety of molecular systems, including proteins, nucleic acids, and small molecules
- OpenMM supports a wide range of force fields and integrators and can run on CPUs and GPUs.
- Open source, written in C++ with Python and other language bindings available

Basic Workflow (object-oriented)

1. Download, complete and edit the structure:
   - **Topology** (i.e. the identity of atoms, bonds, etc)
   - **Positions** (i.e. the starting coordinates)
2. Create the **system** object.
3. Create the **integrator** object.
4. Create and add custom **forces** to system if needed.
5. Define the **simulation** object.
6. Set the initial positions and velocities.
7. Minimize.
8. Run the simulation.
9. (Analyze the results.)

Integrators

- …are algorithms that solve the equations of motion for a system
- OpenMM includes several integrators, e.g. Langevin dynamics, Verlet integrator, and Monte Carlo barostat
- Different integrators are appropriate for different types of simulations and conditions (e.g.: NPT vs NVT)

Simulating a system

- Once a system has been defined and the force field and integrator selected, it can be simulated
- The simulation (run) involves running a series of steps, where each step involves calculating the forces on each atom, integrating the equations of motion, and updating the system's coordinates
- After the simulation, data analysis can be performed to obtain information about the system's behavior and properties

Let’s pick a test system

6H1F: Gelsolin G2+nanobody

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

Display Files | Download Files

Biological Assembly 1

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

Global Symmetry: Asymmetric - C1
Global Stoichiometry: Hetero 2-mer - A1B1

Find Similar Assemblies

Biological assembly 1 assigned by authors and generated by PISA (software)

Biological Assembly Evidence: gel filtration

Macromolecule Content

- Total Structure Weight: 28.49 kDa
- Atom Count: 1,896
- Modelled Residue Count: 229
- Deposited Residue Count: 259
- Unique protein chains: 2

6H1F

Structure of the nanobody-stabilized gelsolin D187N variant (second domain)

PDB DOI: 10.2210/pdb6H1F/pdb

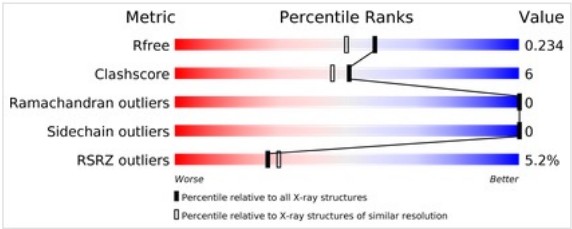
Classification: STRUCTURAL PROTEIN
Organism(s): Lama glama, Homo sapiens
Expression System: Escherichia coli
Mutation(s): Yes

Deposited: 2018-07-11 **Released:** 2019-01-23
Deposition Author(s): Hassan, A., Milani, M., Mastrangelo, E., de Rosa, M.
Funding Organization(s): Amyloidosis Foundation

Experimental Data Snapshot

Method: X-RAY DIFFRACTION
Resolution: 1.90 Å
R-Value Free: 0.233
R-Value Work: 0.199
R-Value Observed: 0.202

wwPDB Validation | 3D Report | Full Report



This is version 1.0 of the entry. See complete history.

Literature | Download Primary Citation

Nanobody interaction unveils structure, dynamics and proteotoxicity of the Finnish-type amyloidogenic gelsolin variant.

Giorgino, T., Mattioni, D., Hassan, A., Milani, M., Mastrangelo, E., Barbiroli, A., Verhelle, A., Gettemans, J., Barzago, M.M., Diomede, L., de Rosa, M.

(2019) Biochim Biophys Acta Mol Basis Dis **1865**: 648-660

PubMed: 30625383 Search on PubMed
DOI: 10.1016/j.bbadis.2019.01.010

Primary Citation of Related Structures:

3D Protein Feature View: 6H1F

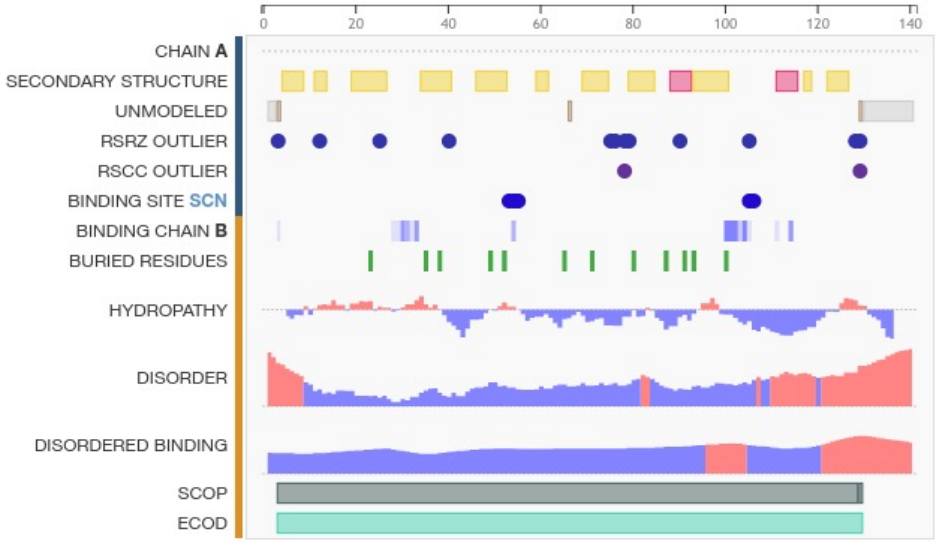
Help Back

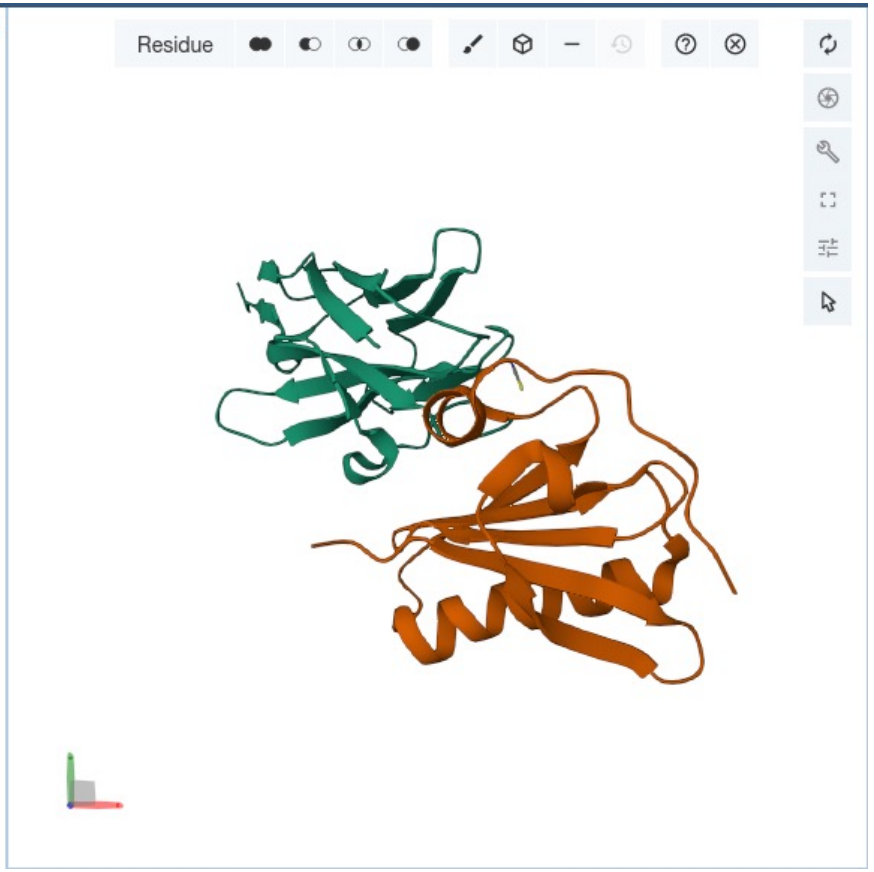
Structure of the nanobody-stabilized gelsolin D187N variant (second domain)

Chain

A

Gelsolin nanobody - Lama glama





Giorgino T, Mattioni D, Hassan A, Milani M, Mastrangelo E, Barbiroli A, et al. Nanobody interaction unveils structure, dynamics and proteotoxicity of the Finnish-type amyloidogenic gelsolin variant. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease. 2019 Mar 1;1865(3):648–60.

Journal link. Preprint.

BBA - Molecular Basis of Disease 1865 (2019) 648–660

Contents lists available at ScienceDirect

BBA - Molecular Basis of Disease

journal homepage: www.elsevier.com/locate/bbadis

Nanobody interaction unveils structure, dynamics and proteotoxicity of the Finnish-type amyloidogenic gelsolin variant

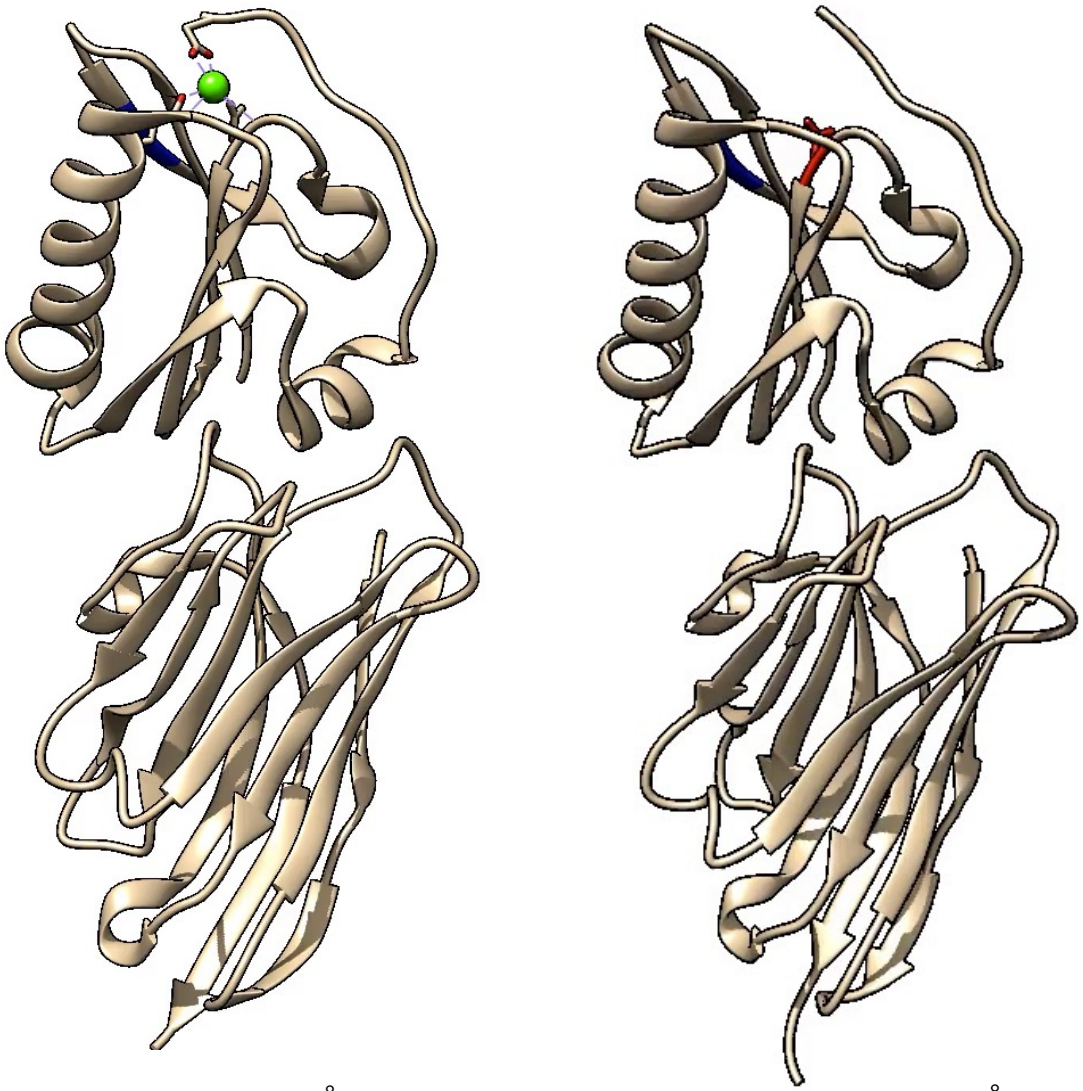
Toni Giorgino[a,b], Davide Mattioni[a,c,1], Amal Hassan[b,1], Mario Milani[a,b], Eloise Mastrangelo[a,b], Alberto Barbiroli[d], Adriaan Verhelle[e], Jan Gettemans[f], Maria Monica Barzago[c], Luisa Diomede[c], Matteo de Rosa[a,b,*]

[a] *Istituto di Biofisica, Consiglio Nazionale delle Ricerche, Milano, Italy*
[b] *Dipartimento di Bioscienze, Università degli Studi di Milano, Milano, Italy*
[c] *Department of Molecular Biochemistry and Pharmacology, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, 20156 Milan, Italy*
[d] *Dipartimento di Scienze per gli Alimenti, la Nutrizione e l'Ambiente, Università degli Studi di Milano, Milano, Italy*
[e] *Department of Molecular Medicine, Department of Molecular and Cellular Neuroscience, Dorris Neuroscience Center, The Scripps Research Institute, La Jolla, CA 92037, USA*
[f] *Nanobody Lab, Department of Biochemistry, Faculty of Medicine and Health Sciences, Ghent University, Ghent, Belgium*

Three puzzles!

WT:Nb11 complex compared to D187N:Nb11.

1. WT and D187N are **virtually identical***: same structure, different function
2. Nb11 binds far from the furin cleavage site…
3. …and far from the Ca^{2+} ion



WT: 4S10, 2.6 Å

D187N: **6H1F**, 1.9 Å

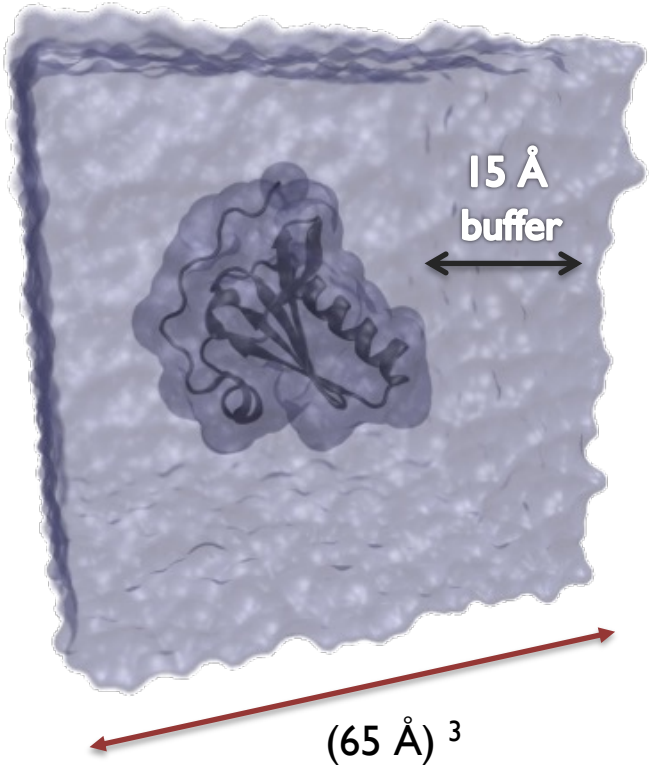
* Except Ca^{2+} binding

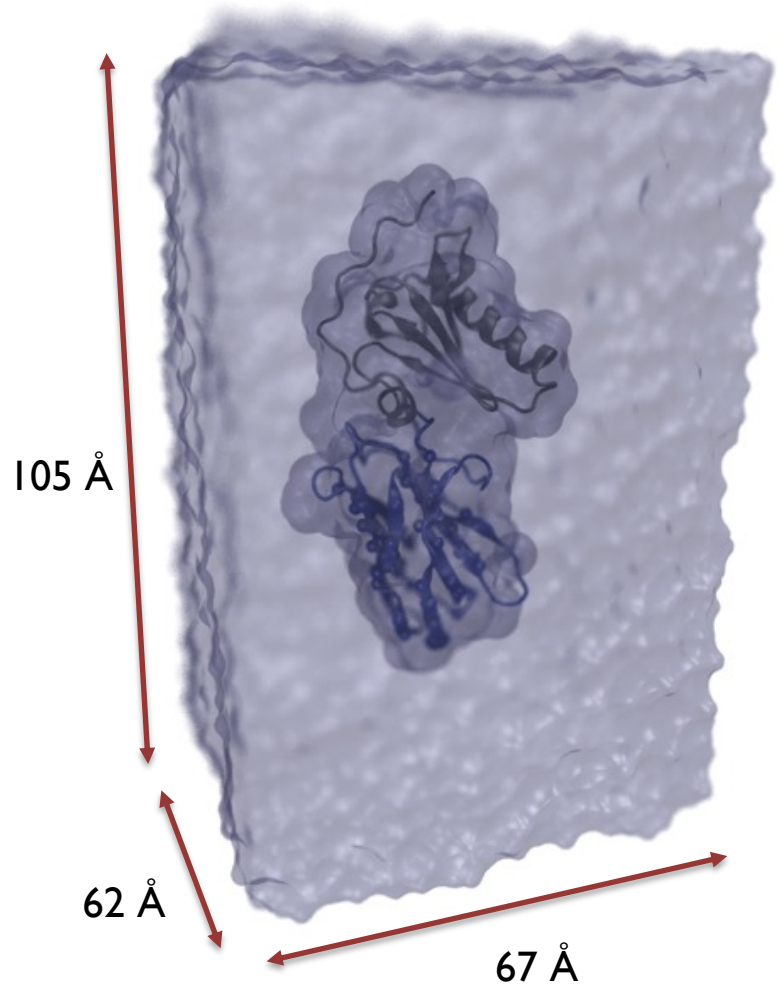
GSN ± Nb11 MD simulations

- Unbiased sampling @300 °K
- 100 mM NaCl
- Harmonic restraints:
  SS Nb1 @ 0.03 kcal/mol/Å^2

CHARMM36

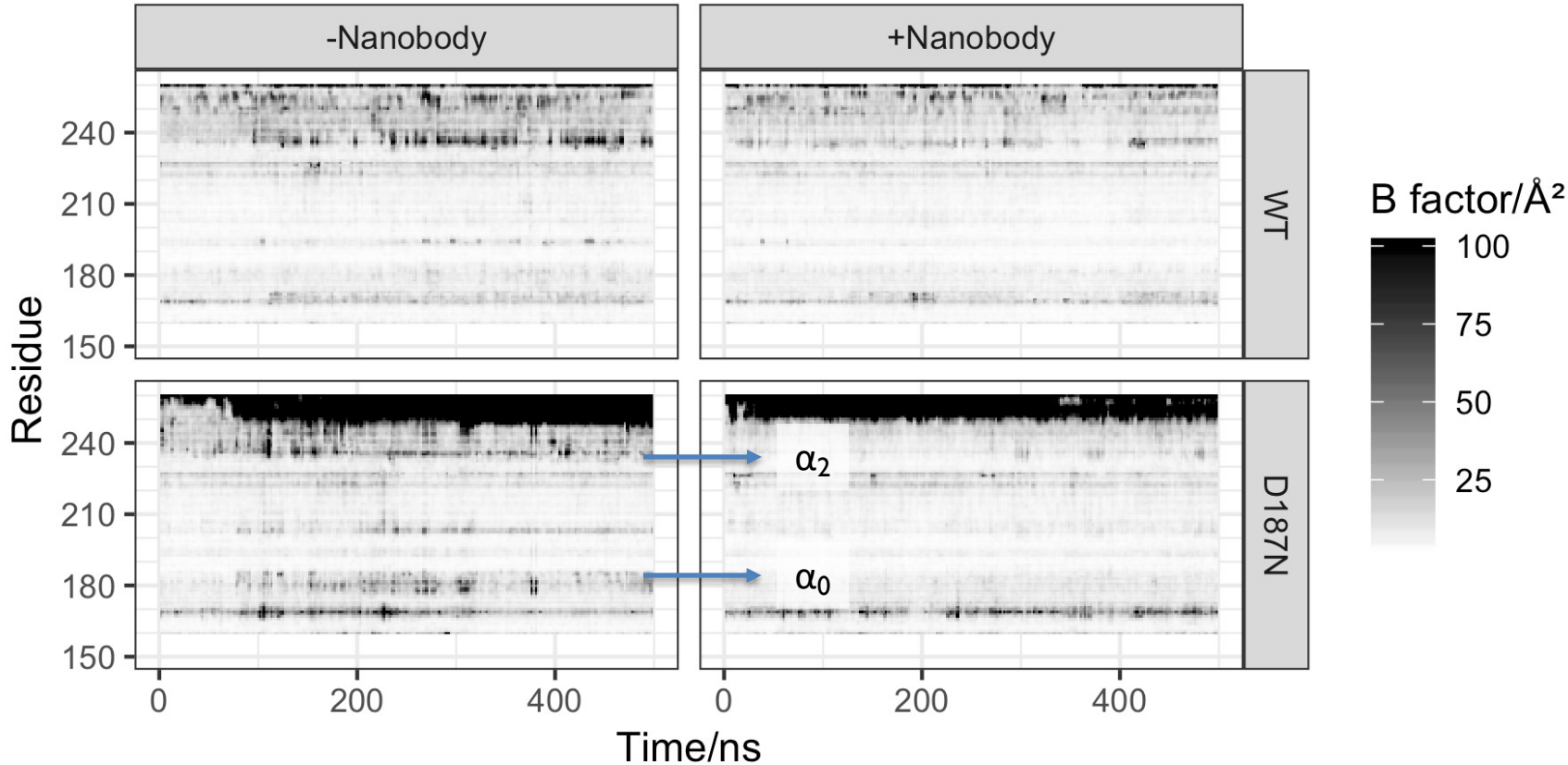
~3 μs tot. ~25k/43k atoms





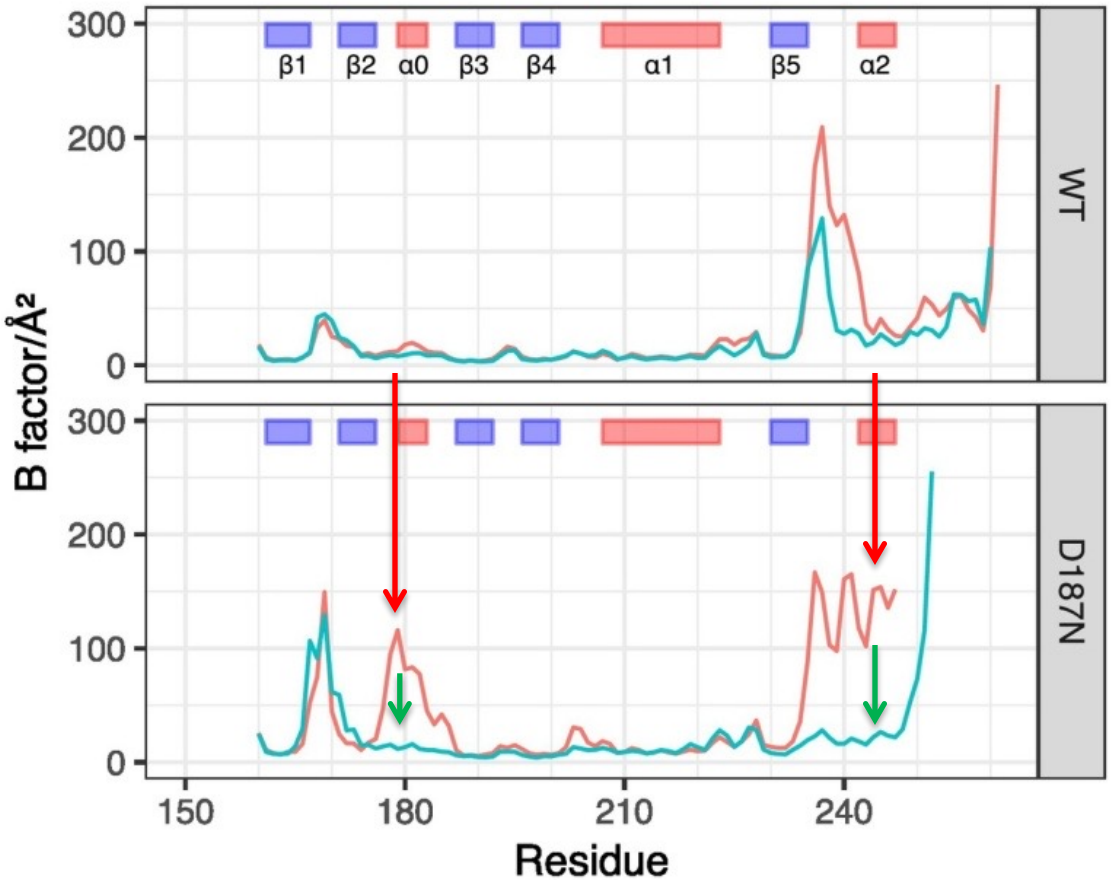
MD results

Sample	Nb11	Ca^{2+}	Simulated time (ns)	C-terminal disorder onset
WT_{G2}	−	+	800	Not observed
WT_{G2}	+	+	750	Not observed
$D187N_{G2}$	−	−	748	After 83 ns
$D187N_{G2}$	+	−	512	After 40 ns



A matter of dynamics?

$B = (8\pi^2/3)\,\mathrm{RMSF}^2$

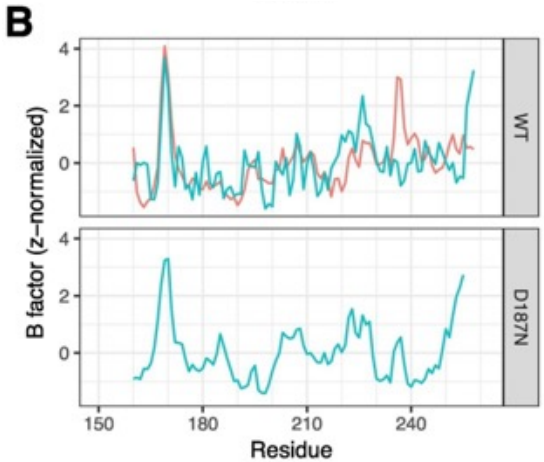


WT: Nb11 has ~ no effect on fluctuation profile (control)

Mutant: destabilized α_0, α_2

Nb11: complete stability recovery

MD vs exp. B-factors →



In practice

Using OpenMM on Google Colab

- We'll test OpenMM on Google Colab to run molecular dynamics simulations without the need for installing any software on your local machine.
- **Google Colab** is a free Jupyter environment that allows you to run Python code in the cloud. GPUs runtimes are available.
- To use OpenMM on Google Colab or locally, open the provided notebook (read the comments)

https://github.com/giorginolab/MD-Tutorial-Data

giorginolab / MD-Tutorial-Data Public

Edit Pins | Unwatch

Code | Issues | Pull requests | Actions | Projects | Wiki | Security | Insights | Settings

main | 1 branch | 0 tags | Go to file | Add file | Code

tonigi Created using Colaboratory 6a4dcdf 7 hours ago 5 commits

GSN	import	3 weeks ago
HIVPR	import	3 weeks ago
notebooks	Created using Colaboratory	7 hours ago
README.md	Initial commit	3 weeks ago

README.md

MD-Tutorial-Data

Data for various MD analysis tutorials

giorginolab / MD-Tutorial-Data Public

Edit Pins

Code | Issues | Pull requests | Actions | Projects | Wiki | Security | Insights | Settings

Code

main

Go to file t

GSN

HIVPR

notebooks

MD-Tutorial-Data / notebooks / 1_OpenMM_build.ipynb

tonigi Created using Colaboratory

Preview Code Blame 573 lines (573 loc) · 15.2 KB

Open in Colab

OpenMM_2023.ipynb

File Modifica Visualizza Inserisci Runtime Strumenti Guida

Condividi

+ Codice + Testo Copia su Drive

Connetti

Colab-specific instructions start here

```
# Here we use a Conda environment inside Google Colab. Blocks specific for Colab
# (like this one) mention "condacolab". On "normal" platforms the procedure
# for installation may be different - you need to check the system's documentation.

# Colab notebooks are "brittle": in the course of time Colab is updated
# and dependencies no longer work properly. Proper HPC platforms are more
# stable (and supported)

# After executing this cell, Colab restarts.

!pip install -q condacolab
import condacolab
condacolab.install_miniforge()
```

```
# Verify Python version
import sys
sys.version
```

```
import condacolab
condacolab.check()
```

```
# Colab-specific workaround for a weird error upon shell escape:
#   NotImplementedError: A UTF-8 locale is required. Got ANSI_X3.4-1968
import locale
def getpreferredencoding(do_setlocale = True):
    return "UTF-8"
locale.getpreferredencoding = getpreferredencoding
```

...when done...

Visualize

- After you have done the simulation, load the minimized PDB and output.dcd in PyMOL
- What about PBCs? Fix with: `pbc_unwrap` …

Questions

- How many atoms?
- How many residues?
- Disulfide bridges?
- How many trajectory frames?
- Simulation length in *actual* time?

More questions

- Does density change? Should it?
- What is the box size? Is it appropriate?
- Relaxation time?
- Plot the log file

Conclusion

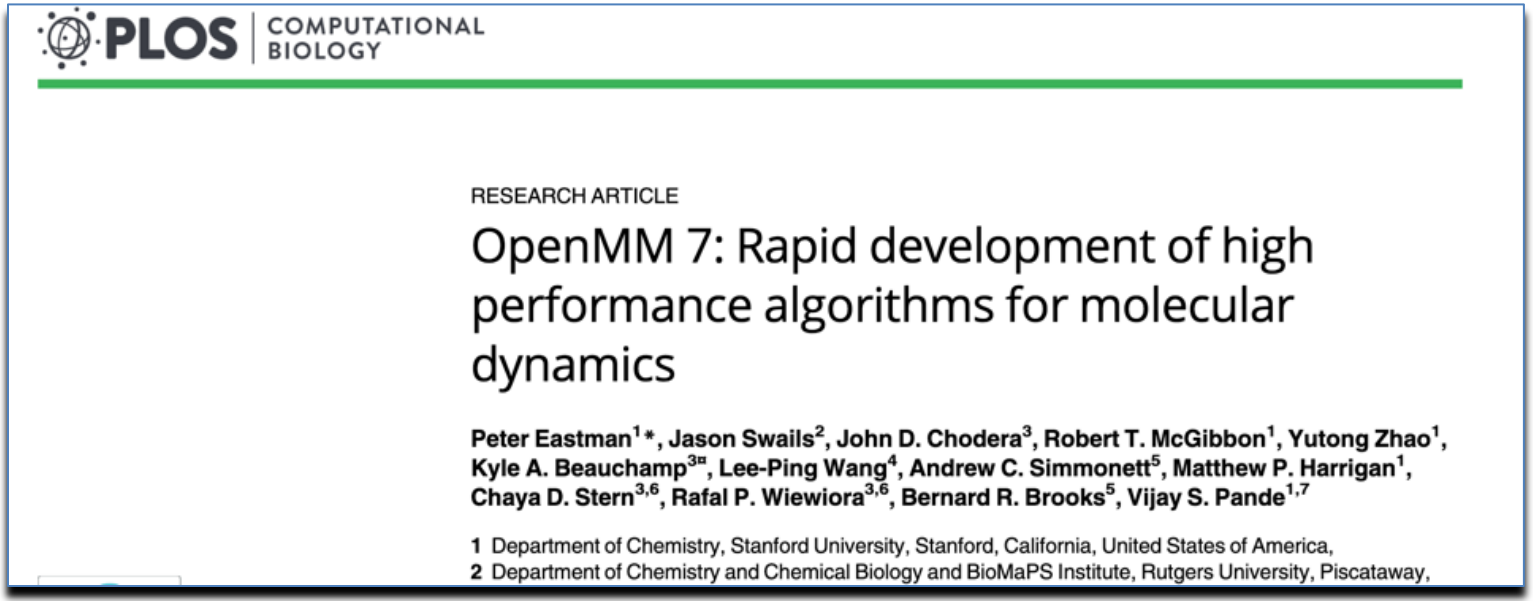
Conclusion

- OpenMM is a powerful tool for molecular dynamics simulations
- Good, if fragmented, documentation
- With its customizable force fields and integrators, it can be used to study a wide range of atomistic systems, e.g.
  - “toy” polymers
  - all-atom MD with major FFs
  - ANN potentials

Resources for learning OpenMM

- OpenMM.org website and documentation
- GitHub repository with examples and tutorials
- Community forums and mailing lists for support and discussion

- See also
  - OpenMMtools
  - https://openforcefield.org/
  - HTMD, ACEMD
  - https://github.com/openmm/pdbfixer
  - Charmm-GUI



PLOS | COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

OpenMM 7: Rapid development of high performance algorithms for molecular dynamics

Peter Eastman[1]*, Jason Swails[2], John D. Chodera[3], Robert T. McGibbon[1], Yutong Zhao[1], Kyle A. Beauchamp[3¤], Lee-Ping Wang[4], Andrew C. Simmonett[5], Matthew P. Harrigan[1], Chaya D. Stern[3,6], Rafal P. Wiewiora[3,6], Bernard R. Brooks[5], Vijay S. Pande[1,7]

1 Department of Chemistry, Stanford University, Stanford, California, United States of America,
2 Department of Chemistry and Chemical Biology and BioMaPS Institute, Rutgers University, Piscataway,

End