

Molecular Dynamics (MD) Simulations in CRISPR-Cas systems

PhD. Pablo Ricardo Arantes

Riverside, January 25th 2023

EDUCATION



2007-2012 Bachelor of Pharmacy
UFRGS – Porto Alegre, BR
Supervisor: Dr. Hugo Verli

2013-2014 M. Sc. in Cellular &
Molecular Biology
PPGBCM - UFRGS – Porto Alegre, BR
Supervisor: Dr. Hugo Verli

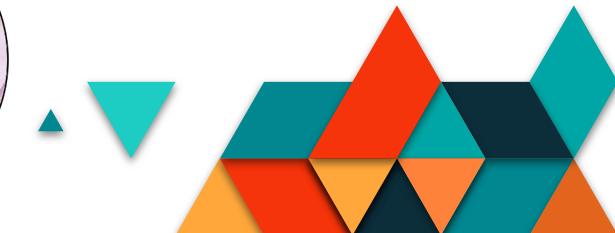
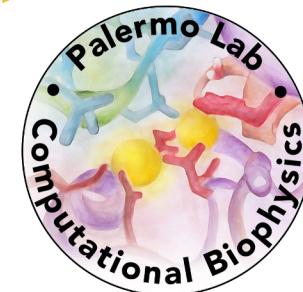
2014-2018 PhD in Cellular &
Molecular Biology
PPGBCM - UFRGS – Porto Alegre, BR
Supervisor: Dr. Hugo Verli

2018-2019 Postdoctoral Researcher
PPGBio - UFCSPA - Porto Alegre, BR
Supervisor: Dr. Dinara Jaqueline Moura

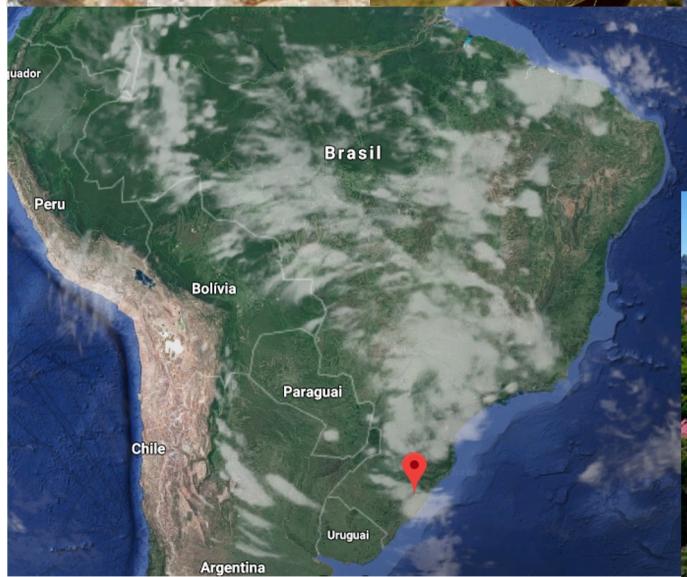
2019-2023 Postdoctoral Researcher
University of California, Riverside
Supervisor: Dr. Giulia Palermo



Universidade Federal de Ciências da Saúde
de Porto Alegre



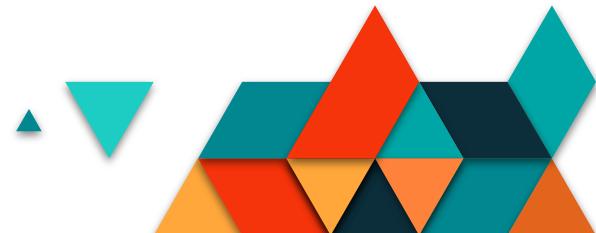
Porto Alegre - Brazil



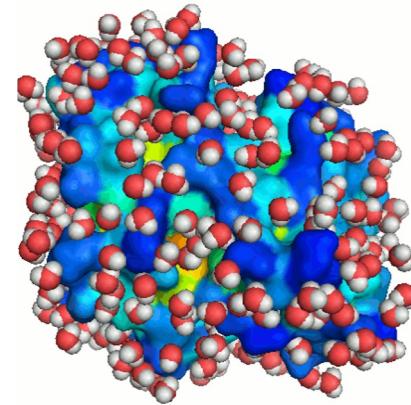
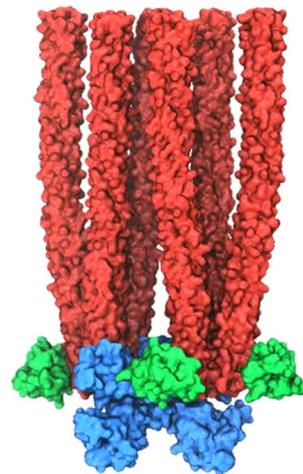
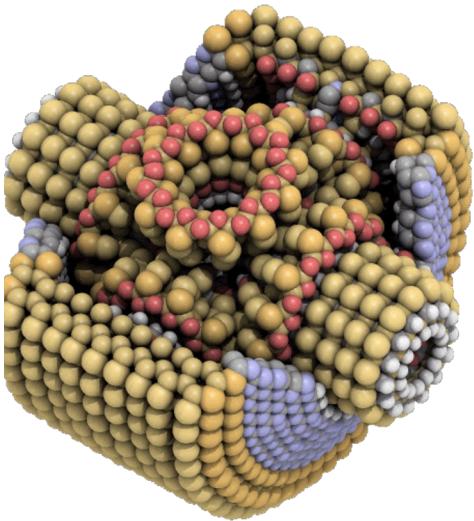
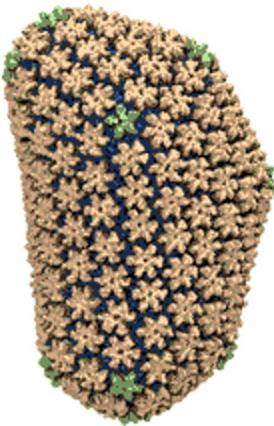
Slides

Jupyter Notebook Example

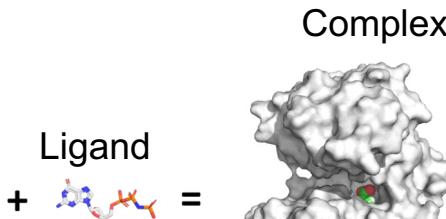
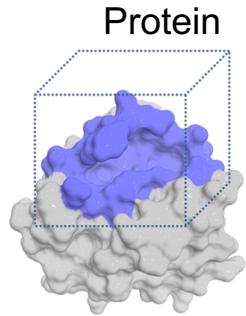
<https://github.com/pablo-arantes/BIEN-249>



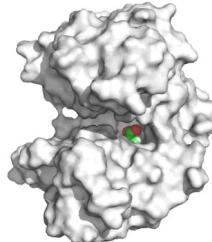
MOLECULAR MODELING?



COMPUTATIONAL CHEMISTRY

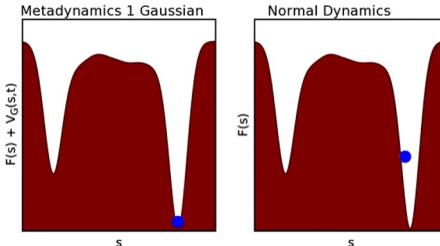
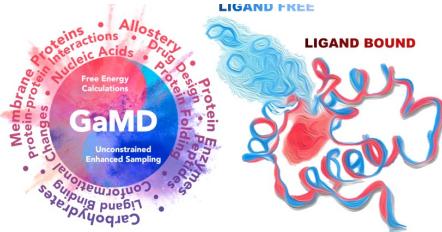
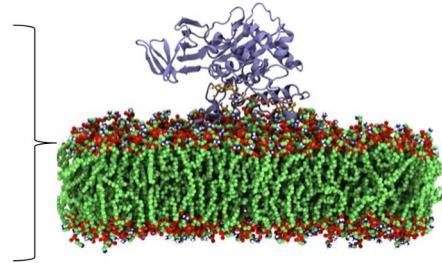


Complex



Acceleration
of atom i

$$\frac{d^2\mathbf{r}_i(t)}{dt^2} = \frac{\mathbf{F}_i}{m_i} \rightarrow \begin{array}{l} \text{Force on atom } i \\ \text{Mass of atom } i \end{array}$$



MOLECULAR MODELING

Programs: AlphaFold2, Modeller; Rosetta; I-Tasser; Phyre 2;

MOLECULAR DOCKING - CHEMIOINFORMATICS

Programs: Autodock Vina; Autodock; DockThor (Brazil); RDKit; TorchANI

MOLECULAR DYNAMICS (Molecular Mechanics)

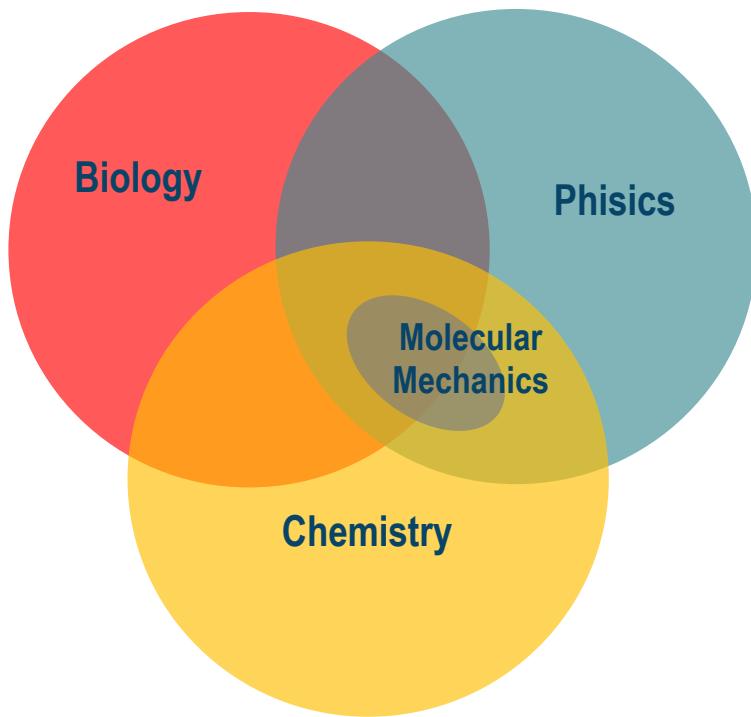
Programs: Amber; OpenMM; GROMACS; NAMD;

ENHANCED SAMPLING

Programs: Amber; GROMACS; PLUMED; GaMD;



MOLECULAR DYNAMICS (Molecular Mechanics)



What is Molecular Dynamics?

MD is a technique that allows us to study the dynamic movement of atoms and molecules.

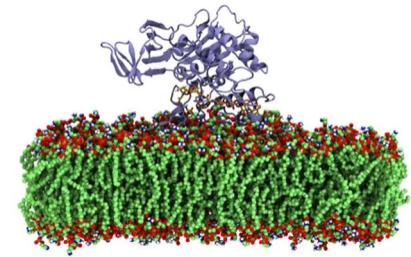
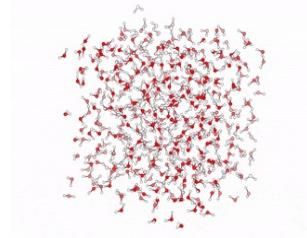
“Molecular Dynamics (MD) is a computer simulation technique where the time evolution of a set of interacting atoms is followed by integrating their **equations of motion**.” [1]

Through integration, you obtain the trajectories of each atom in the system. The trajectory files gives the position and velocity information for each atom; allowing you to track the movement of each atom.

Through use of analytical techniques (RMSD, PCA, RMSF, Cluster Analysis, etc.) we can thoroughly study conformational variability of macromolecules (proteins and nucleic acids).

$$\frac{d^2\mathbf{r}_i(t)}{dt^2} = \frac{\mathbf{F}_i}{m_i}$$

Acceleration of atom i → Force on atom i
Mass of atom i →

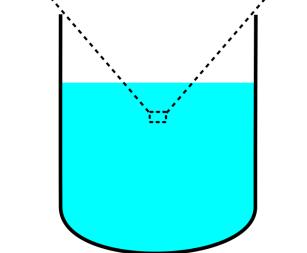
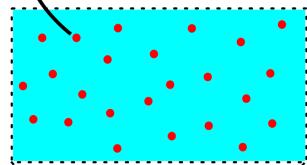
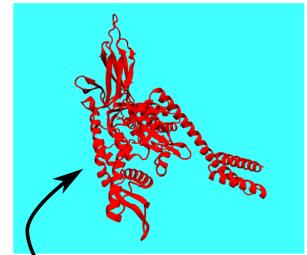


Ergodic hypothesis

In most cases, MD simulations are performed for a single replica of a macromolecule of interest. On the contrary, in real solutions there are multiple molecules present.

Ergodic hypothesis assumes that the time evolution of a single molecule may be used to describe the thermodynamic properties of an ensemble of molecules.

$$\langle A \rangle_{ensemble} = \langle A \rangle_{time}$$



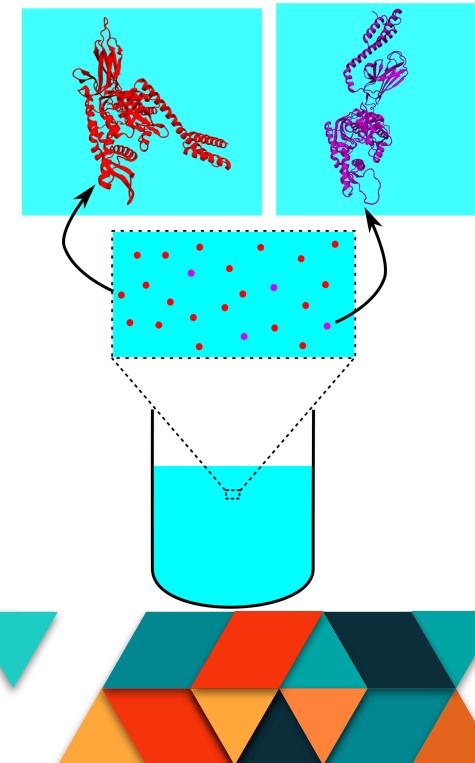
Connections between MD and experiments

The equilibrium constant in experiments quantify the ratio of the macromolecules in two states , e.g. open (red) and closed (magenta)

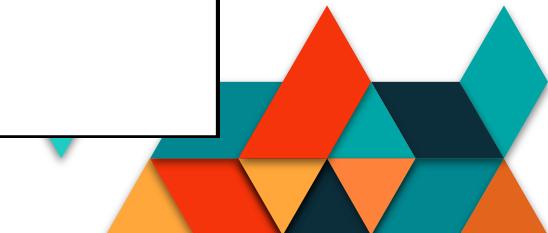
$$K_{exp} = \frac{N_{open}}{N_{closed}}$$

In Molecular Dynamics simulations, the equilibrium constant quantify the ratio of the probabilities for the macromolecule to occupy two states

$$K_{MD} = \frac{p_{open}}{p_{closed}}$$



Connections between MD and experiments



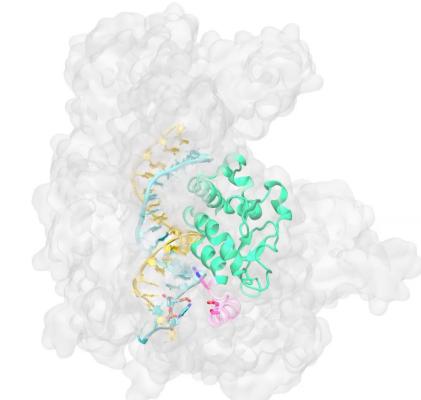
Current Applications

MD is being used to perform experiments that would otherwise be impossible or very expensive to do in reality.

MD simulations is often used as a preliminary step to real life experiments. This can results in saved time and money for experiments that might be unnecessary.

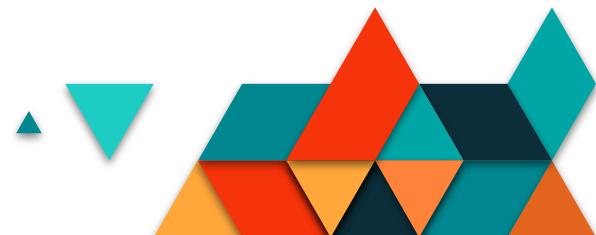
MD being currently being used to study Drug Design, Material Science, Macromolecules etc.

The Palermo Lab is specifically using MD to study the CRISPR – Cas9 systems.

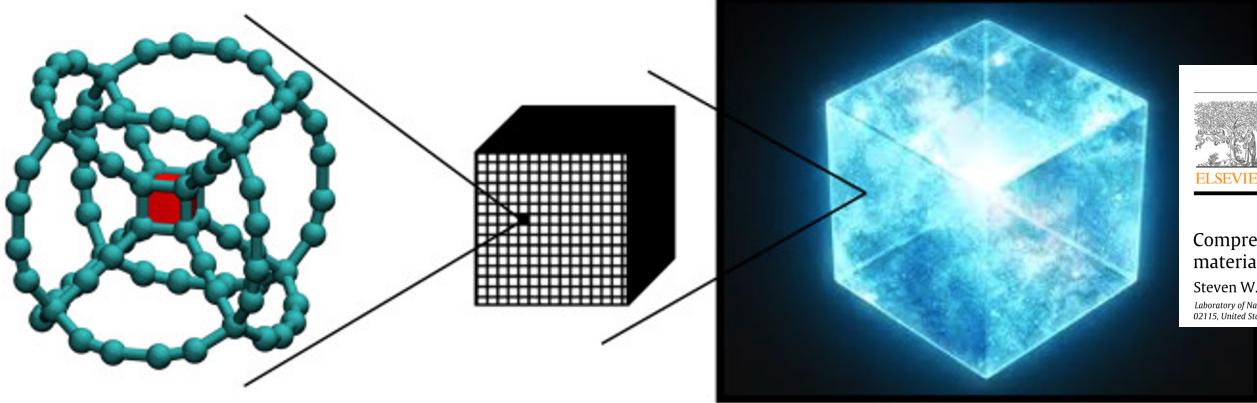


CRISPR-Cas9

**Base pair mismatches
“lock” HNH in the
inactive state
(“checkpoint”)**



Current Applications



Extreme Mechanics Letters 22 (2018) 19–26

Contents lists available at ScienceDirect



Extreme Mechanics Letters

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



Compressive failure of a carbon nano-tesseract: Sci-Fi inspired materials and the strength of thanos

Steven W. Cranford

Laboratory of Nanotechnology In Civil Engineering (NICE), Department of Civil and Environmental Engineering, Northeastern University, Boston, MA 02115, United States



40000
tons
750000 x
strength of a human



MD is continually being improved

Todays simulations far more complicated, since we can create larger systems with todays more powerful hardware. This further extends how Molecular Dynamics can be applied. Parameters that describe our system are also continually being improved make today's MD simulations more realistic.

Longer MD trajectories (increased number of time steps)

Larger Systems (increased number of atoms)

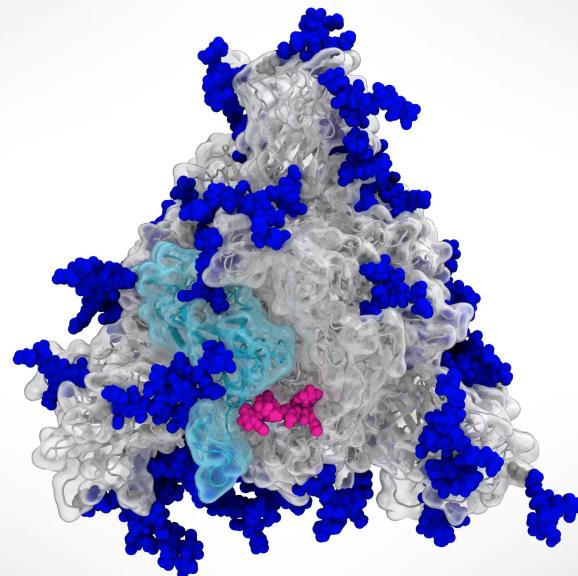
Higher Quality Force Field Parameters (Key for future development)



MD is continually being improved

Closed Spike

Top view



MOVIE BY L. CASALINO

T. SZTAIN, S.-H. AHN et al.
AMARO LAB (UCSD)
CHONG LAB (PITT)

Supercomputing-driven simulations depict the glycan N343 (magenta) acting as a molecular crowbar to pry open the SARS-CoV-2 spike's receptor binding domain, or RBD (cyan), from a "down" to an "up" position. **Nature Chemistry, 2021.**

MD is continually being improved

ACS
central
science



This is an open access article published under an ACS AuthorChoice [License](#), which permits copying and redistribution of the article or any adaptations for non-commercial purposes.



ACS
AUTHORCHOICE

FIRST REACTIONS

Fighting COVID-19 Using Molecular Dynamics Simulations



Cite This: *ACS Cent. Sci.* 2020, 6, 1654–1656



Read Online

ACCESS |

 Metrics & More

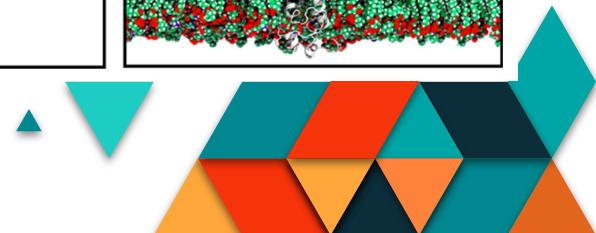
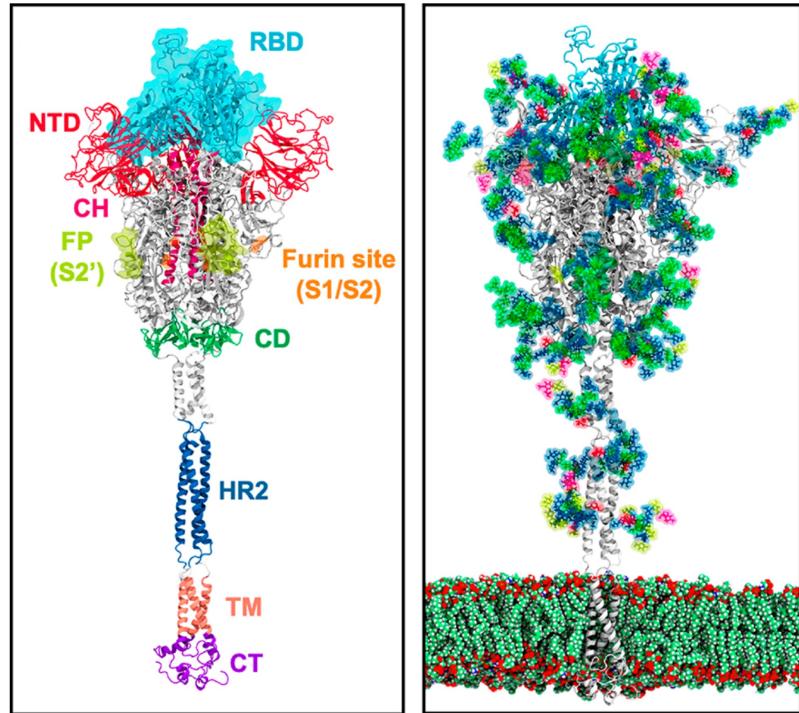
 Article Recommendations

Pablo R. Arantes, Aakash Saha and Giulia Palermo



MD is continually being improved

Molecular dynamics simulations revealed a promising immune target on the SARS-CoV-2 spike protein, proposing novel strategies for vaccine development.



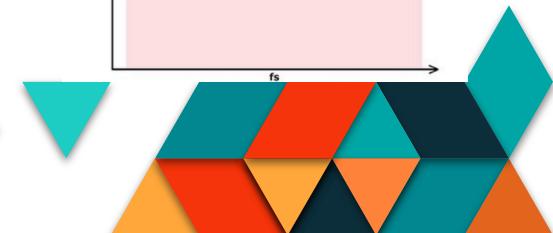
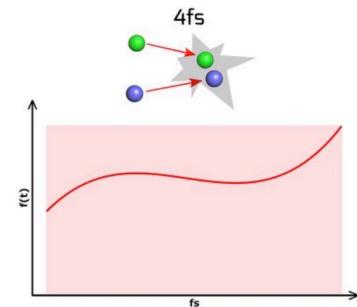
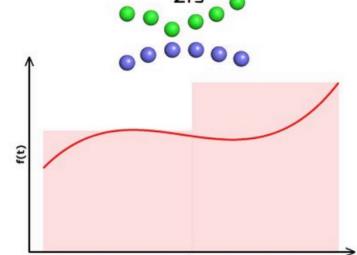
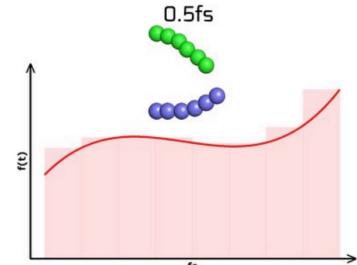
Limitations and Constraints

One major limitation is access to computational power. Decisions need to be made on the:

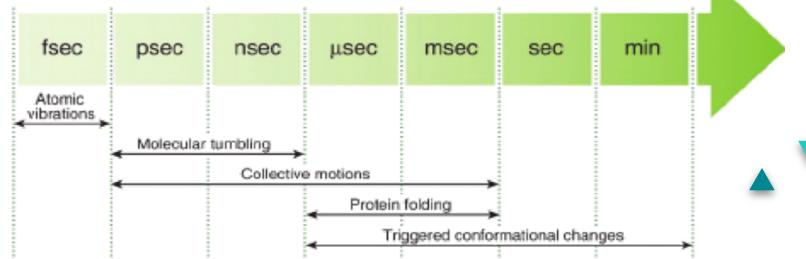
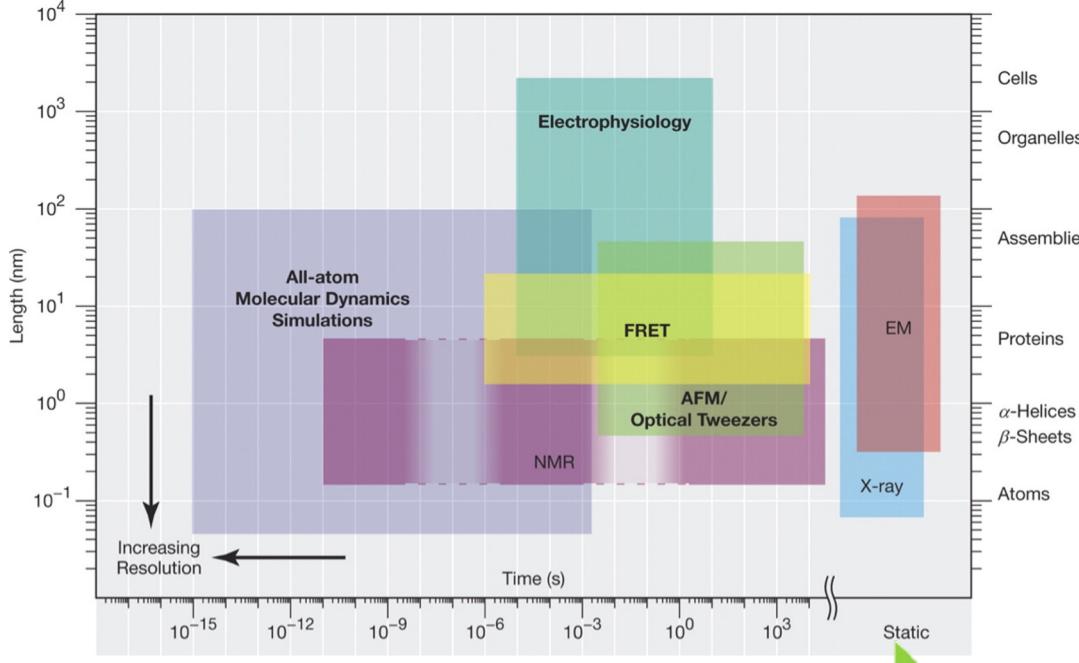
- **Size** of your system
- **Duration** of your system
- **# of timesteps** of your system

Bonds cannot be broken with standard classical molecular simulations. More computational heavy **Quantum Molecular (QM)** simulations would be required to simulate the breakage of a bond.

In MD simulations, one of the more computational heavy step involves determining the potential of all of the atoms within your system. **Particle-Mesh Ewald Summation** can be applied to alleviate some of the computational demands of this step.



Limitations and Constraints



MD Simulation Accuracy

Force Field Selection

- The incorrect selection will result in the atoms moving in an incorrect manner

Duration

- During a MD simulation, a large number of states will be sampled. The duration of the simulation needs to be long enough to sample all of the states.

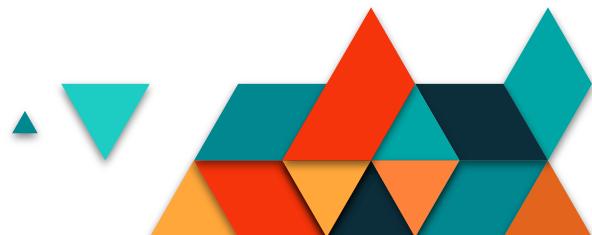
Box size

- The size of the simulation box needs to be large enough to avoid observing boundary condition artifacts.

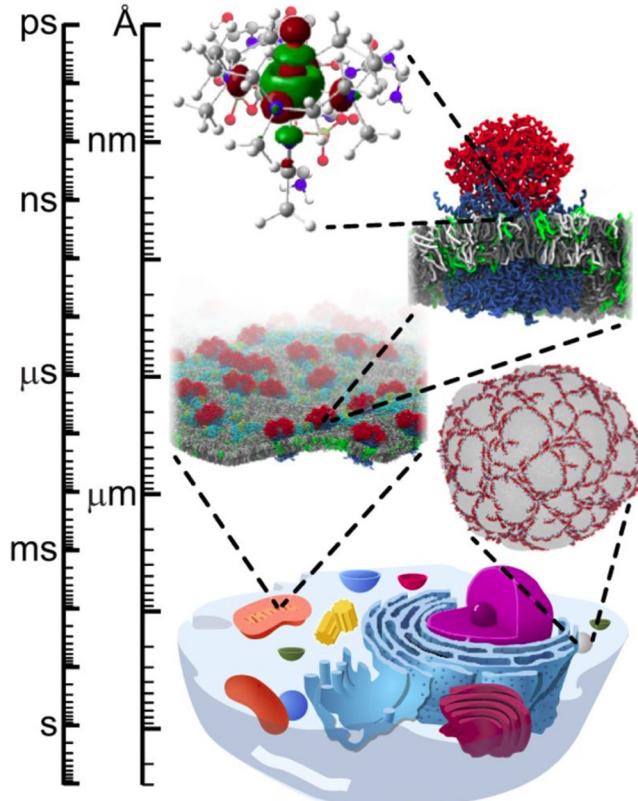
Time-step size

The size of your time-step needs to be determined by the highest frequency bond in your system

The “SHAKE” Algorithm can be used to fix the high frequency bonds to allow for a longer time-step



MD Simulation Accuracy



Quantum

- atoms, electrons and electron clouds included
- explicit solvent
- quantum mechanics

All-atom

- all or most atoms present
- explicit solvent
- molecular dynamics

Coarse-grained

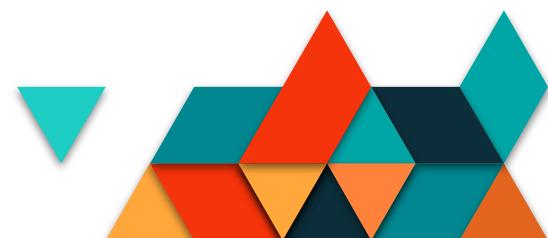
- beads comprising a few atoms
- explicit or implicit solvent
- molecular dynamics

Supra-coarse-grained

- interaction sites comprising many atoms, protein parts or proteins
- implicit solvent
- stochastic dynamics

Continuum

- materials as a continuous mass
- implicit solvent
- continuum mechanics



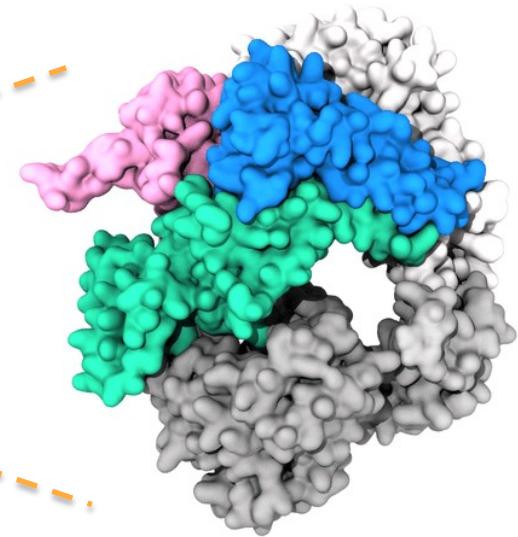
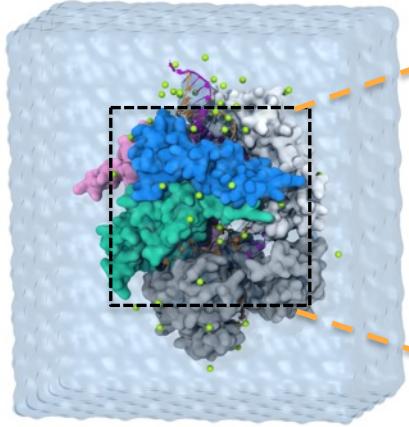
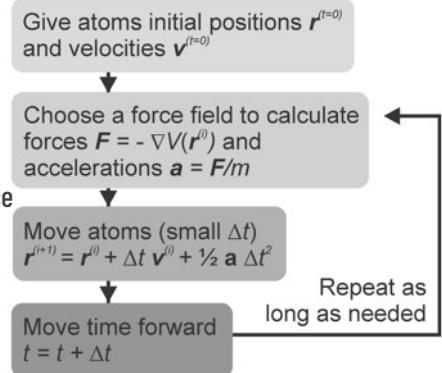
Molecular Dynamics Cycle

Molecular Dynamics (MD)

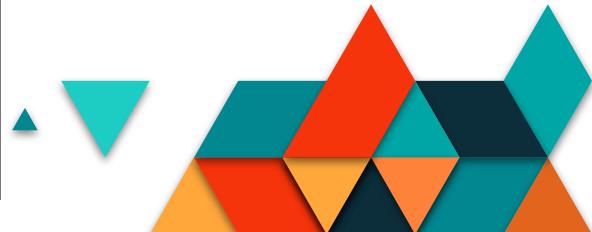
3D-structures:

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

- Cryo-electron microscopy
- X-ray crystallography
- Nuclear magnetic resonance spectroscopy (NMR)
- AlphaFold2

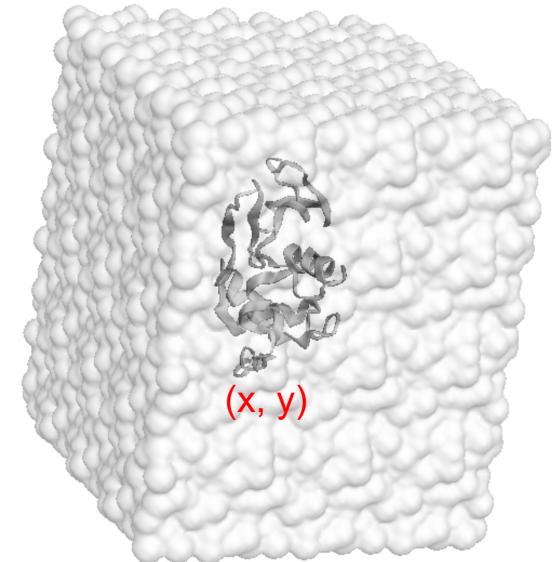


Anton-2



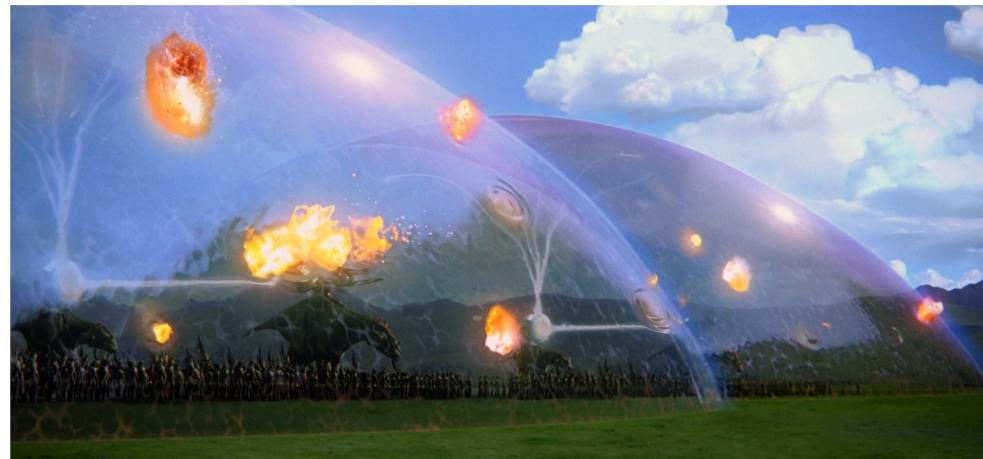
Define the Initial State

- **# of atoms (N) in the system**
- You need to know in **initial position** and the **initial velocity** at time zero for all the atoms.
- Assign the **positions** of the atoms
- Assign the **velocities** of the atoms
- The collection of all of the atoms at a given time is called a the initial configuration.



Application of the Force Field

- The force field will describe the interactions between particles. **The force field is a potential energy function consisting of bonded and nonbonded terms that describe all of the interactions.**
- This is a critical step. The incorrect selection of the force field will result in either description of the interactions that is not detailed enough or the simulation length that is insufficient to sample the desired conformational transition.

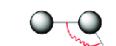


$$U(R) = \sum_{bonds} k_r (r - r_{eq})^2 + \sum_{angles} k_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} k_\phi (1 + \cos[n\phi - \gamma]) + \sum_{impropers} k_o (\omega - \omega_{eq})^2 + \sum_{atoms} \epsilon_{ij} \left[\left(\frac{r_m}{r_{ij}} \right)^{12} - 2 \left(\frac{r_m}{r_{ij}} \right)^6 \right] + \sum_{i,j} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}$$

bond



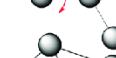
angle



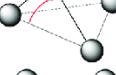
dihedral



improper



van der Waals



electrostatic



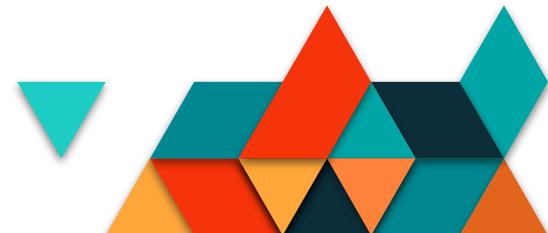
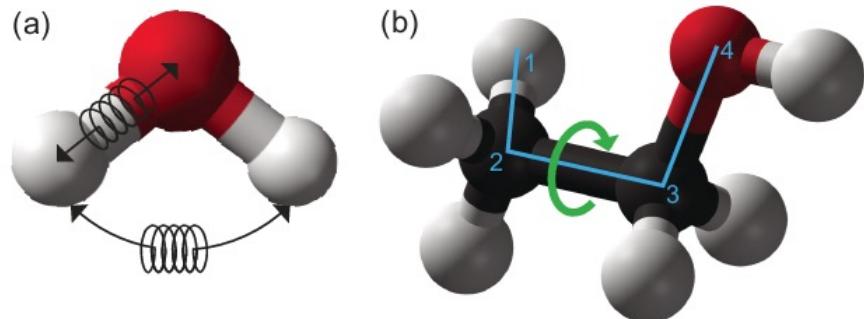
Application of the Force Field

- **Bonded terms:**

Stretching terms, angle bending terms, terms describing the rotation of torsional angles.
Explained by using Harmonic Potentials

- **Non-Bonded terms:**

Coulomb electrostatic interactions
Van der Waals interactions



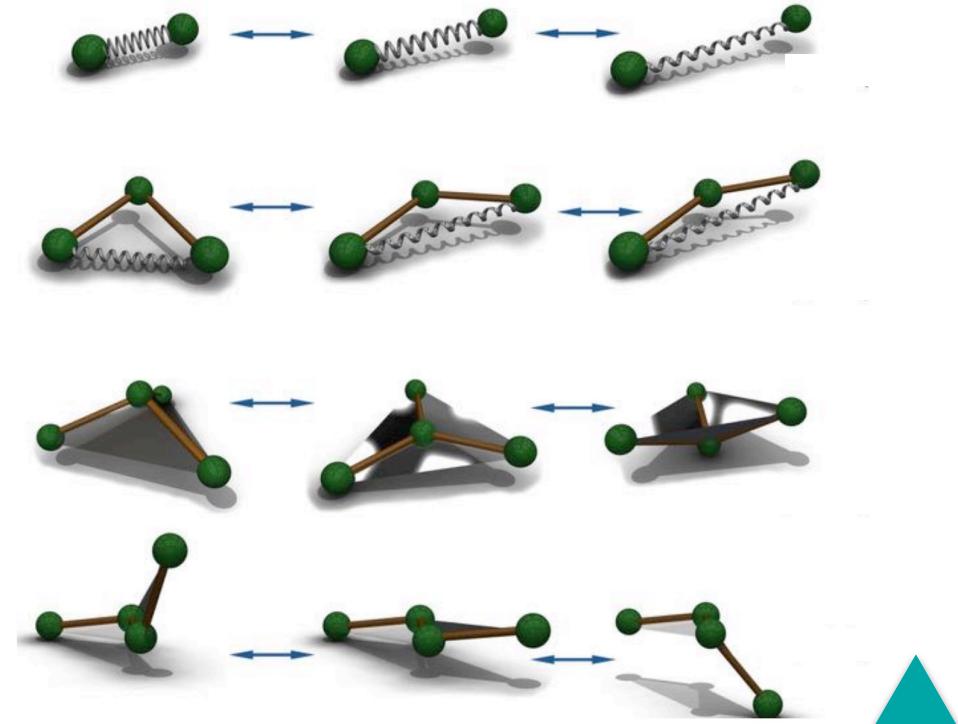
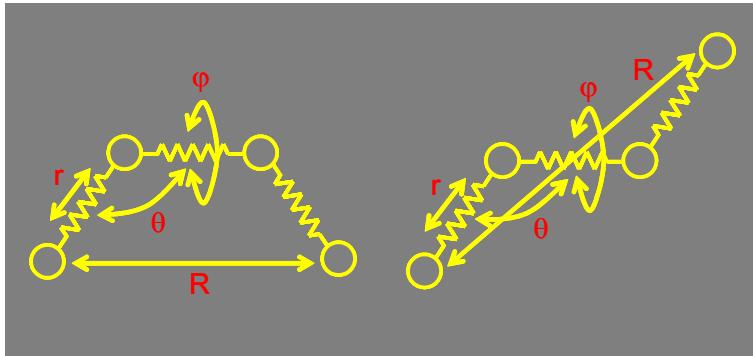
Application of the Force Field

$$U(\text{forcefield})$$

$$= \sum_{\text{bonds}} k_b (r - r_0)^2$$

$$+ \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2$$

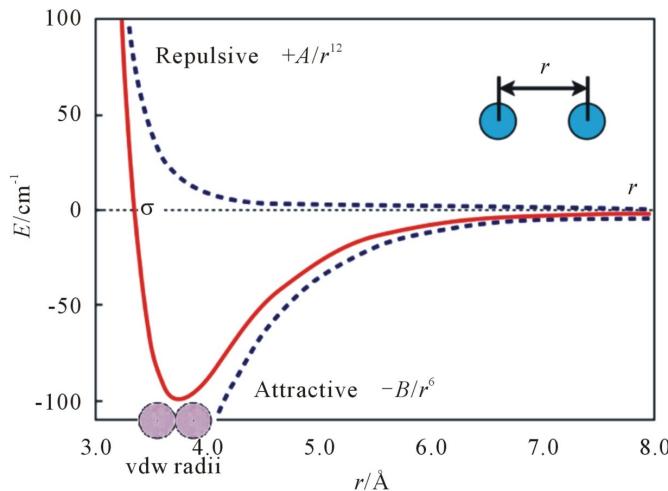
$$+ \sum_{\text{torsions}} A [1 + \cos(n\varphi - \gamma)]$$



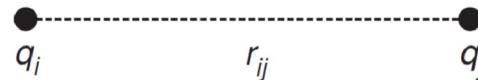
Application of the Force Field

$$\left. \begin{aligned} &+ \sum_{vdW} \left(-\frac{A}{r^6} + \frac{B}{r^{12}} \right) \\ &+ \sum_{Coulomb} \frac{q_1 q_2}{\epsilon r} \end{aligned} \right\} \text{Non-bonded}$$

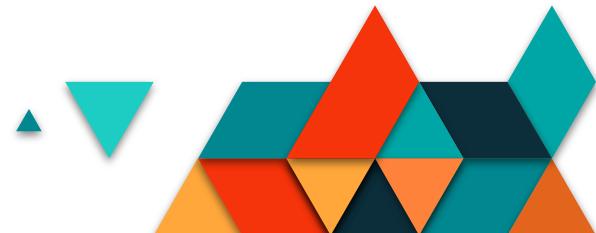
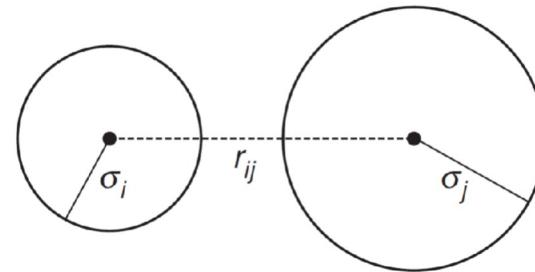
Lennard – Jones Potential



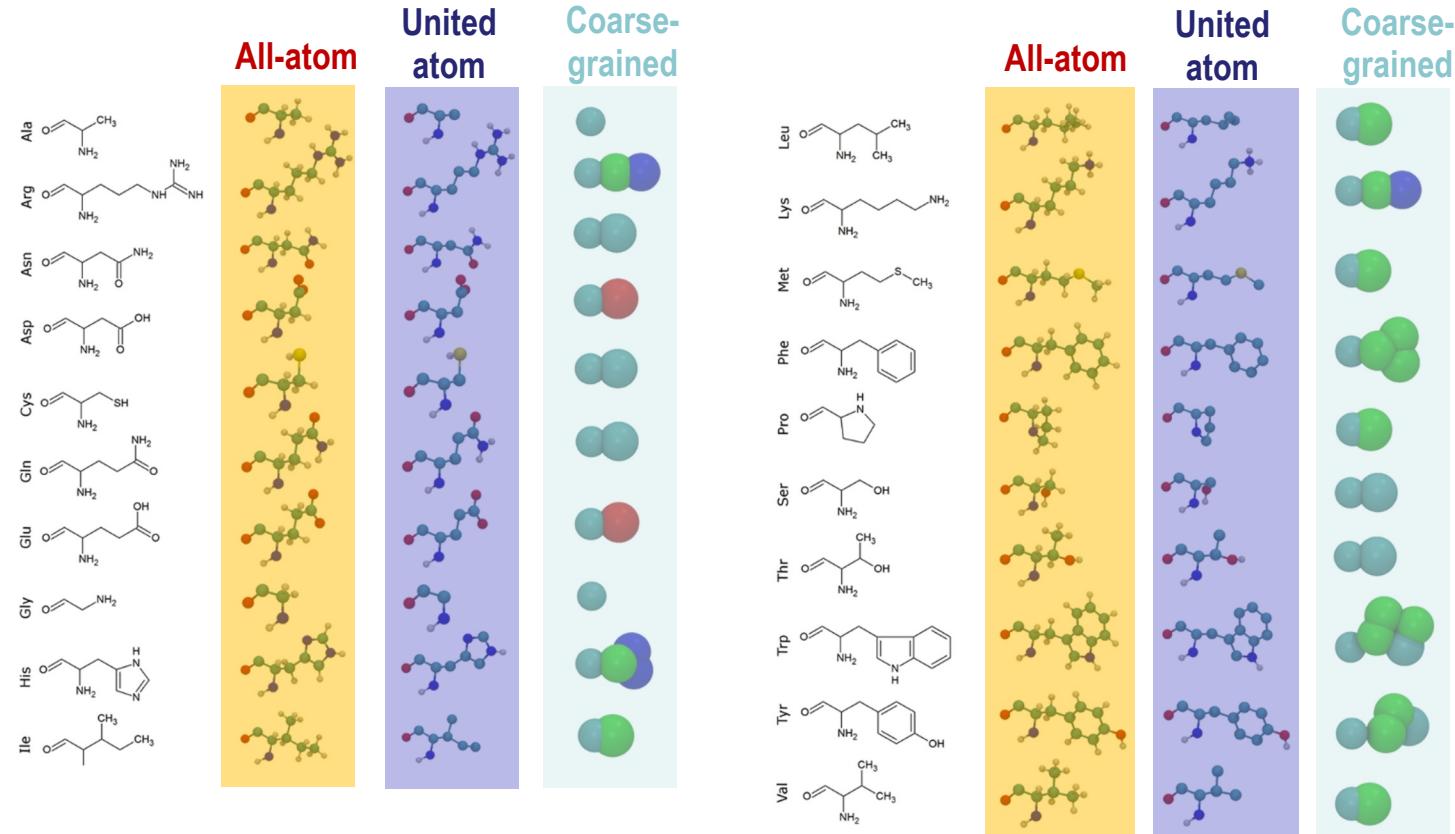
Coulomb electrostatic interactions



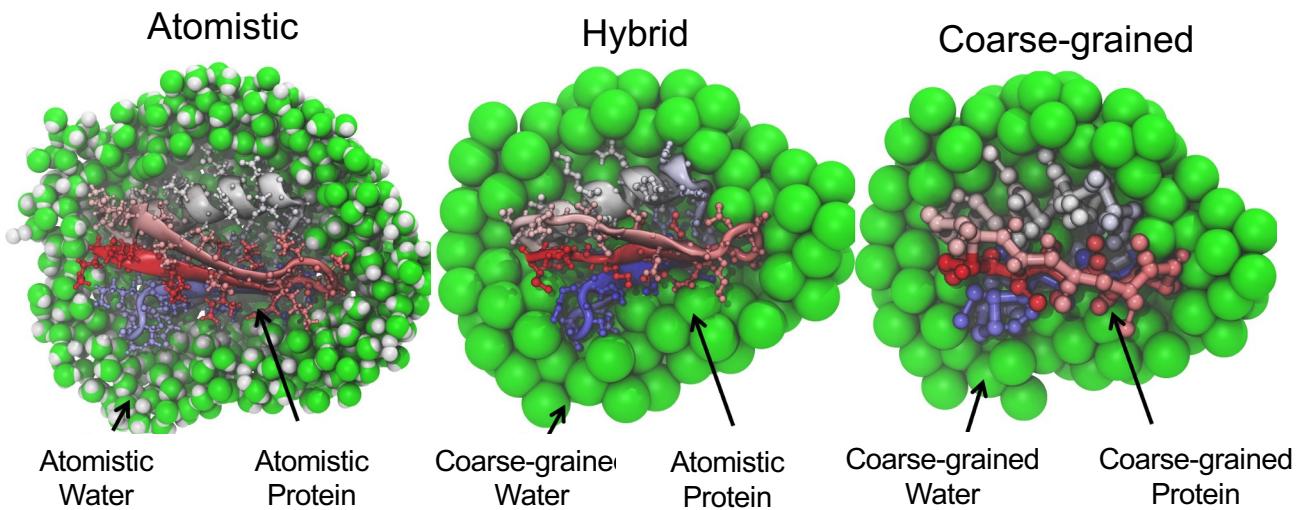
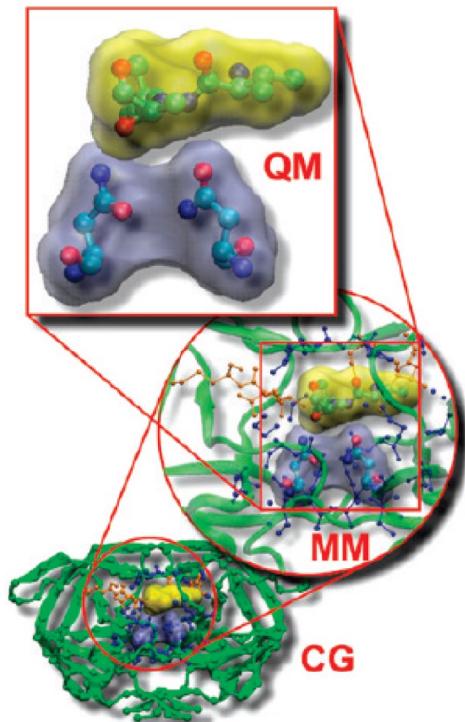
Van der Waals interactions



Application of the Force Field



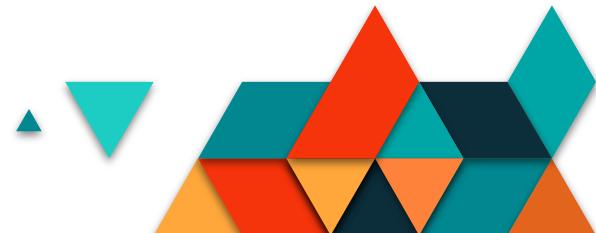
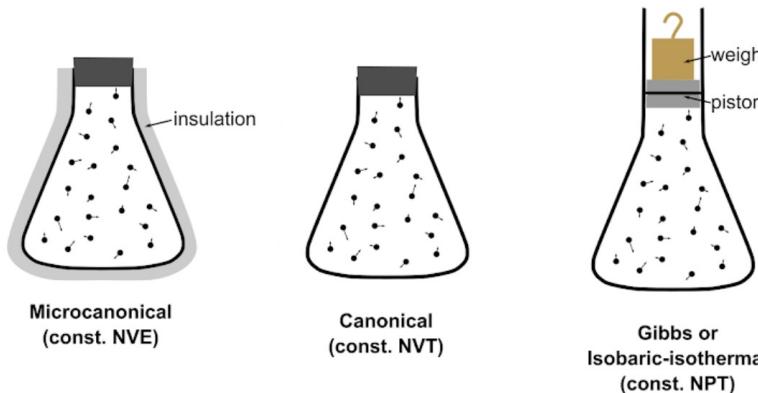
Application of the Force Field



MD Ensembles

You will have to decide the ensemble that you want your simulation to run in:

- **Microcanonical ensemble (NVE):** Simulation of system based on constant-number of particles (N), constant-volume (V), and constant-energy (E);
- **Canonical Ensemble (NVT):** Simulation of system based on constant number of particles (N), constant-volume (V), and constant-temperature (T);
- **Isothermal-isobaric ensemble (NPT):** Simulation of system based on constant number of particles (N), constant-pressure (P), and constant-temperature (T);
St:



Total Energy

- Total energy is the sum of the potential and kinetic energies.
- The Total Potential energy is the sum of all of the pairwise interactions for that particular configuration.

$$\text{Total Energy: } E = K + U$$

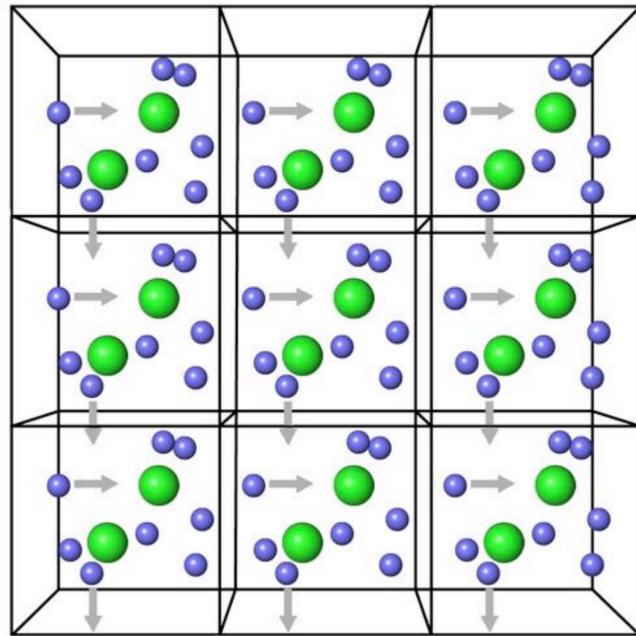
$$\text{Total Kinetic Energy: } K = \sum_i \frac{1}{2} m_i \mathbf{v}_i^2$$

$$\text{Total Potential Energy: } U(\mathbf{r}_1, \dots, \mathbf{r}_N) = \sum_i \sum_{j>i} \phi(|\mathbf{r}_i - \mathbf{r}_j|)$$



Periodic Boundary Conditions (PBC)

- The simulated system is a sub-ensemble within an infinite system of identical, small ensembles. The solute is replicated in all surrounding periodic images, being exactly 1 in each box.
- PBC approximates bulk properties from simulations.
- “PBC does not effectively simulate an infinitely sized simulation box, though they do reduce many otherwise egregious finite – size effects” [2]





The Nobel Prize in Chemistry 2013

The Nobel Prize in Chemistry 2013

Martin Karplus, Michael Levitt, Arieh Warshel



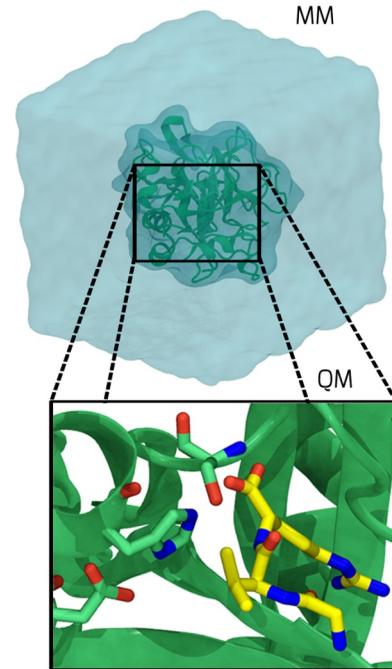
Martin Karplus



Michael Levitt



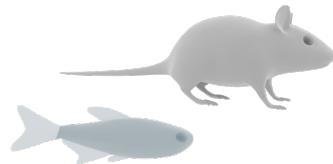
Arieh Warshel



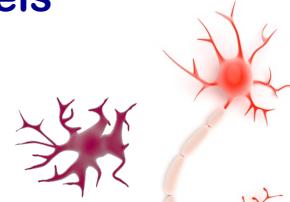
CRISPR-Cas systems

The Gene Editing Revolution

Biology



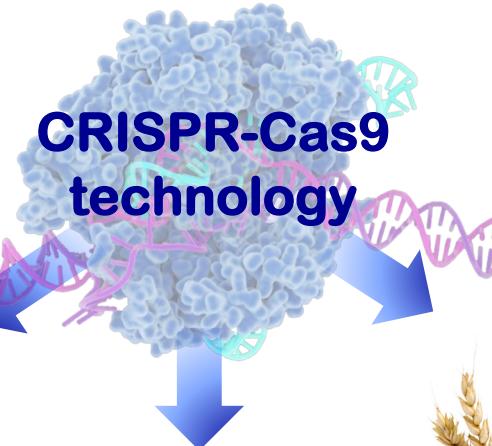
Animal
models



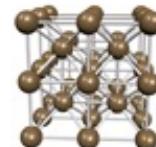
Gene
variation

Medicine

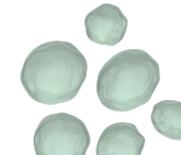
Gene surgery



Materials



Fuels



Food



The Nobel Prize in Chemistry 2020
awarded "for the development of a
method of genome editing"



Biotech

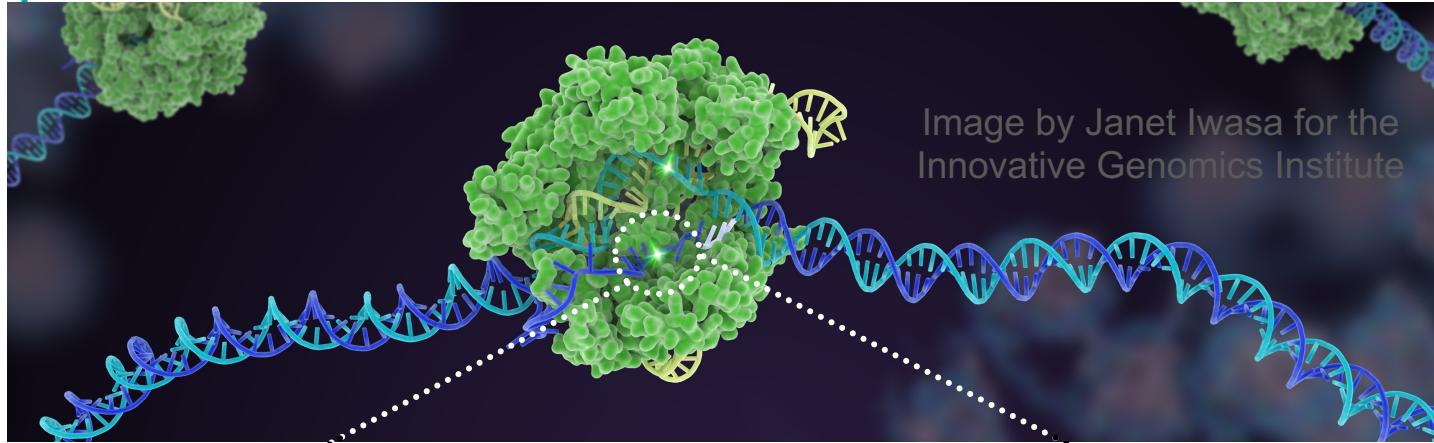


Drug discovery



CRISPR-Cas systems

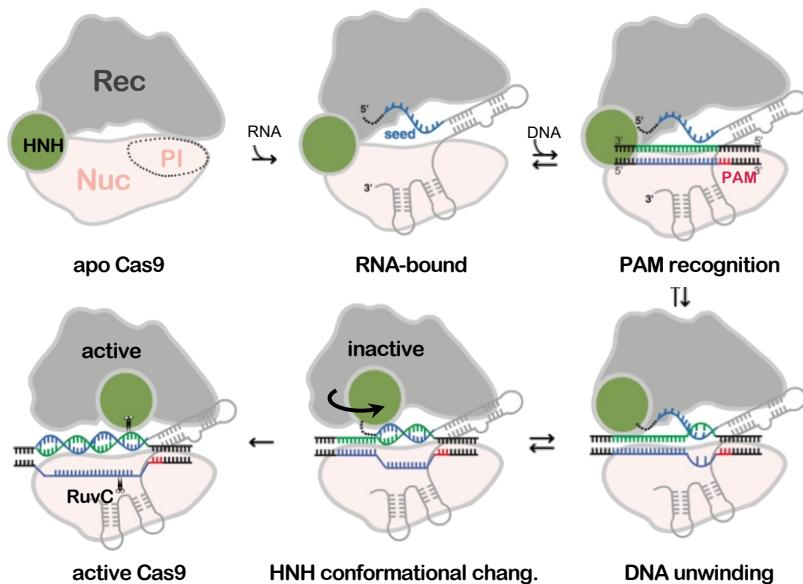
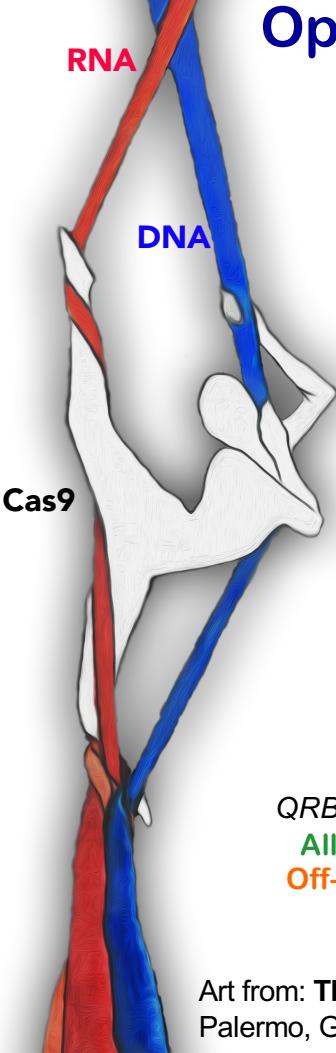
Anton2: world's fastest supercomputer for MD



Microsoft Azure



Open Questions for Molecular Dynamics



Jiang F. & Doudna J. A. *Annu. Rev. Biophys.* 2017

Conformational activation: PNAS 2017; ACS Cent Sci 2016; QRB 2018; Structure 2019; Front. Mol. Biosci. 2021; WIREs Comp. Sci. 2021

Allosteric signaling: JACS 2017; JACS 2020 & WIREs Comp. Sci. 2020

Off-target effects: ACS Cent Sci 2019, Chem 2020, Front Mol Biosci 2020

Catalysis: JCIM 2019 & ACS Catalysis 2020

Art from: **The Invisible Dance of CRISPR-Cas9.**

Palermo, G., Ricci, C., McCammon J.A. *Physics Today*, April 2019



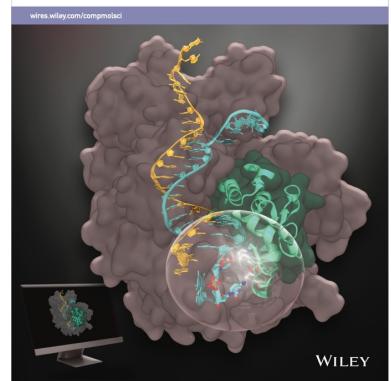
ACS Publications

Most Trusted. Most Cited. Most Read.

Saha and Arantes et al. *JCIM* 2020

Volume 11, Number 3, May/June 2021

WIREs
COMPUTATIONAL MOLECULAR SCIENCE

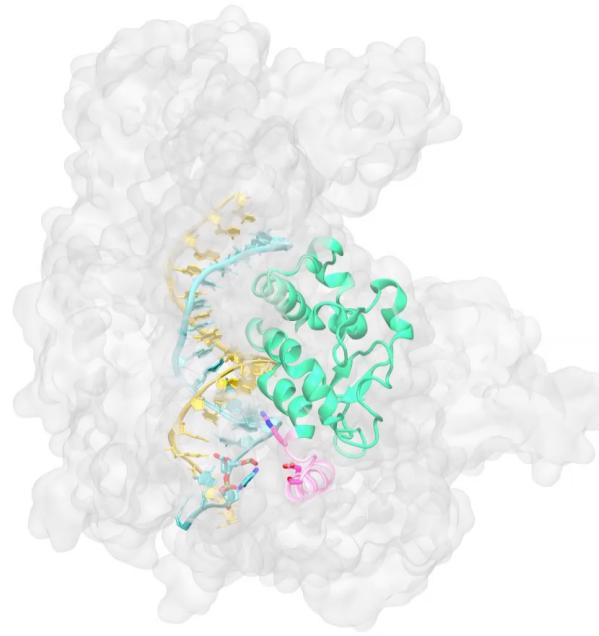


Nierwizcki and Arantes et al. *WIREs Comp. Sci.* 2020

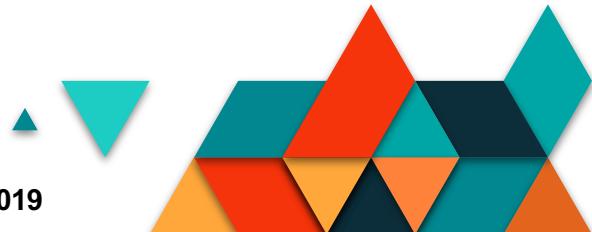
CRISPR-Cas systems

Results from MD simulations

Base pair mismatches
“lock” HNH in the
inactive state
(“checkpoint”)

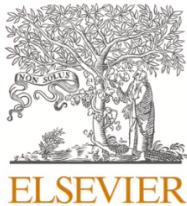


CRISPR-Cas9



CRISPR-Cas systems

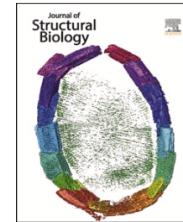
Journal of Structural Biology 214 (2022) 107814



Contents lists available at [ScienceDirect](#)

Journal of Structural Biology

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



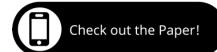
Structural and dynamic insights into the HNH nuclease of divergent Cas9 species

Helen B. Belato^a, Alexandra M. D'Ordine^a, Lukasz Nierzwicki^b, Pablo R. Arantes^b, Gerwald Jogl^a, Giulia Palermo^{b,c,*}, George P. Lisi^{a,*}

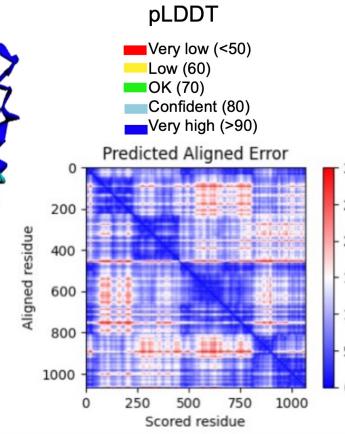
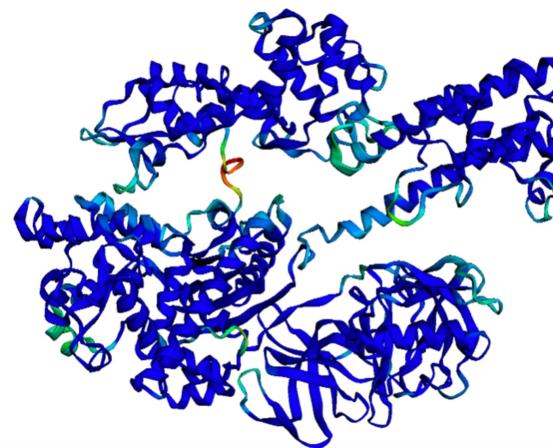
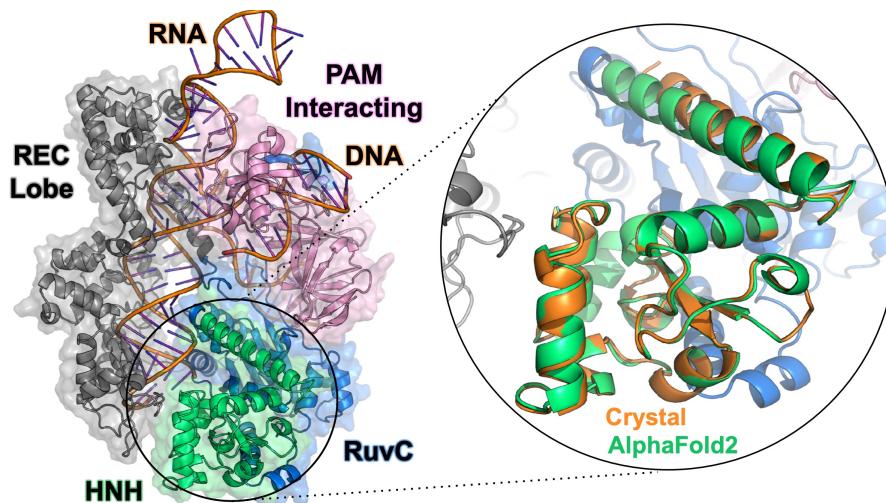
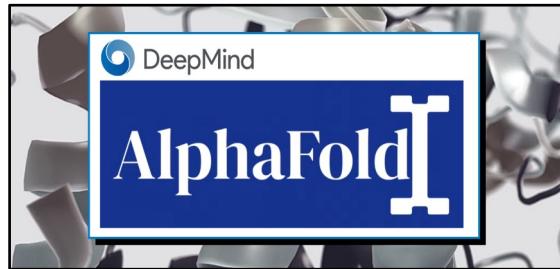
^a Department of Molecular Biology, Cell Biology & Biochemistry, Brown University, Providence, RI, USA

^b Department of Bioengineering, University of California Riverside, Riverside, CA, USA

^c Department of Chemistry, University of California Riverside, Riverside, CA, USA



CRISPR-Cas systems



CRISPR-Cas systems



Emerging Methods and Applications to Decrypt Allostery in Proteins and Nucleic Acids



Pablo R. Arantes, Amun C. Patel and Giulia Palermo *

Department of Bioengineering, University of California Riverside, 900 University Avenue, Riverside, CA 92512, United States

Department of Chemistry, University of California Riverside, 900 University Avenue, Riverside, CA 92512, United States

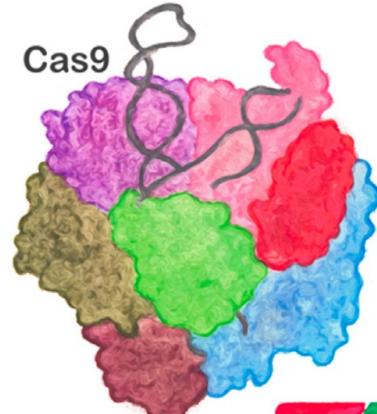
Correspondence to Giulia Palermo: giulia.palermo@ucr.edu (G. Palermo), [@pablitoarantes](https://twitter.com/pablitoarantes) (P.R. Arantes), [@palermo_lab](https://twitter.com/palermo_lab) (G. Palermo)

<https://doi.org/10.1016/j.jmb.2022.167518>

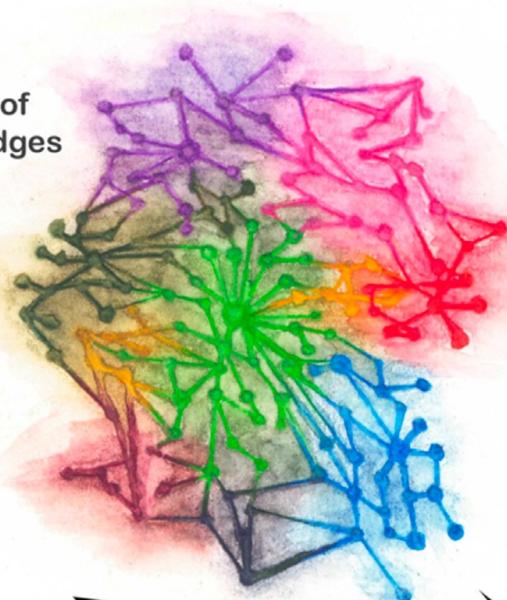
Edited by Igor Berezovsky

CRISPR-Cas systems

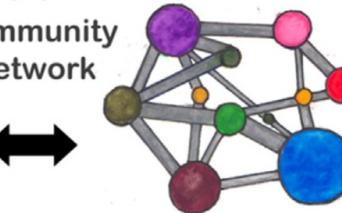
(a) biomolecule



(b)
network of
nodes & edges



(c)
community
network



(d)

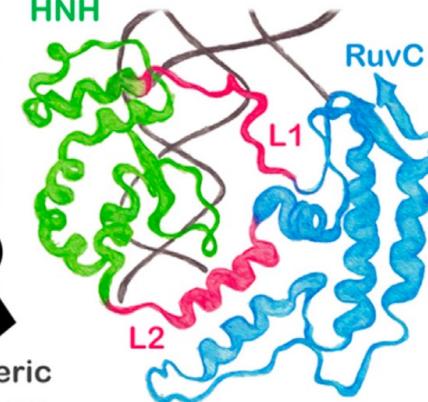
GC



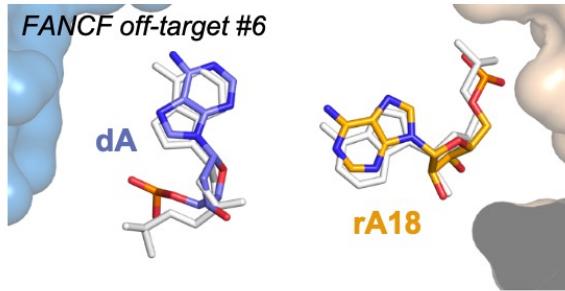
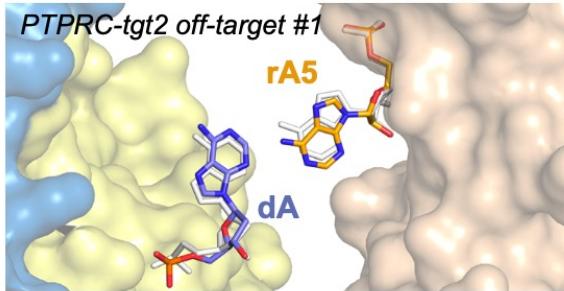
$$w_{ij} = -\log GC$$

(e)

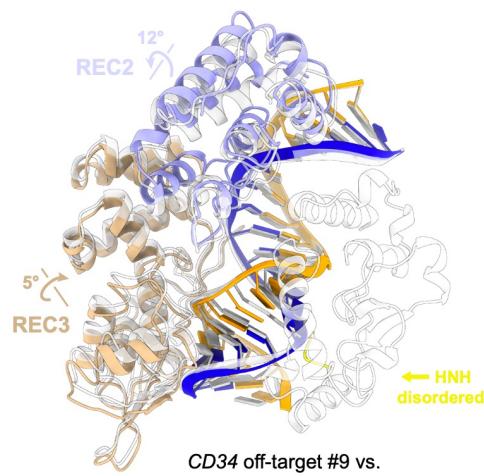
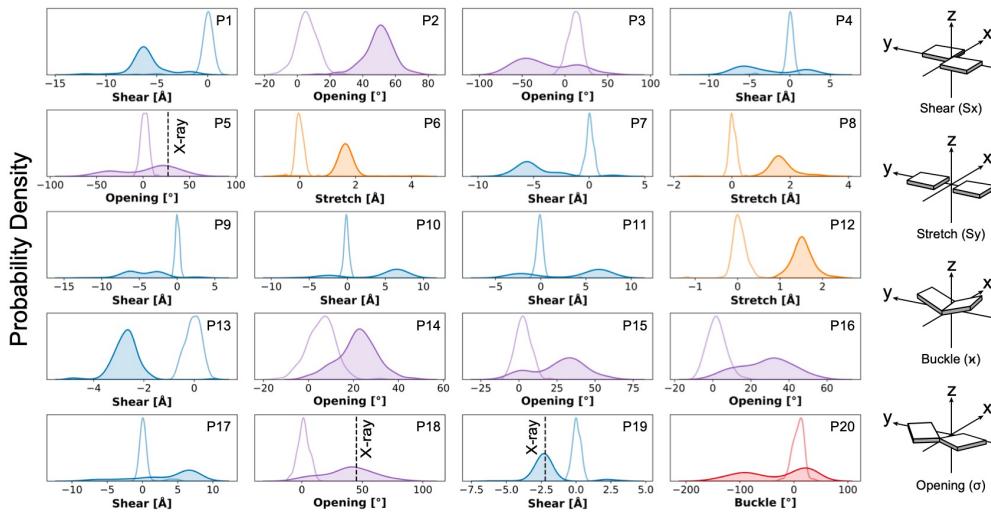
allosteric
pathway



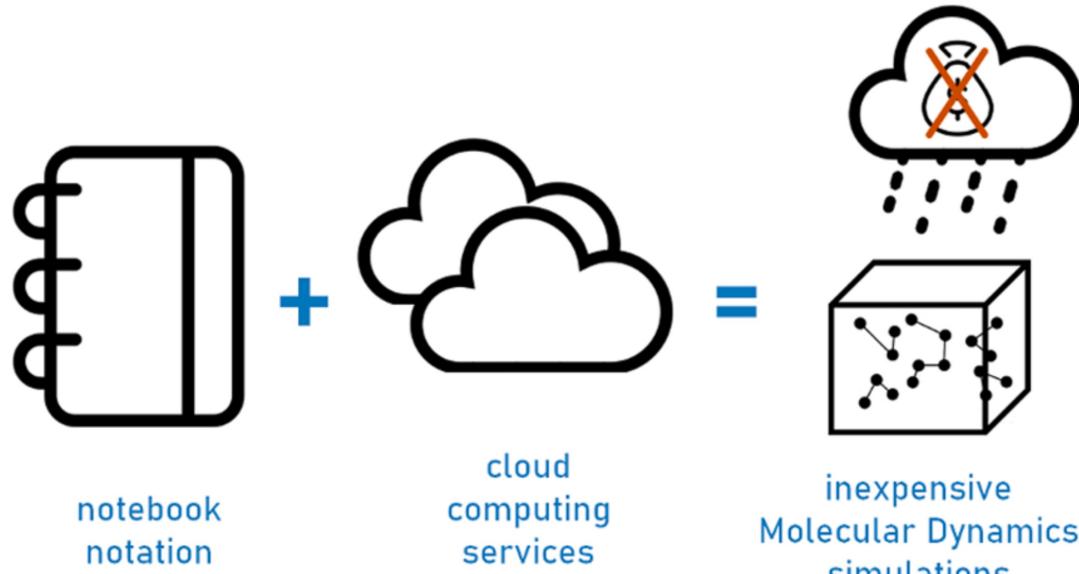
CRISPR-Cas systems



Prof. Martin Jinek
@University of Zürich



Steps for MD run



<https://pablo-arantes.github.io/making-it-rain/>



Steps for MD run

1. System preparation
2. Minimization/Relaxation
3. Equilibration
4. Production Run

▶ Install dependencies

► It will take a few minutes, please, drink a coffee and wait. ;-)

[Show code](#)



<https://github.com/pablo-arantes/BIEN-249> ▾



1: System Preparation

- Build the Configuration of the system (unfavorable configuration should be avoided)
- Build the Starting Structure
- Solvating (when necessary)
- Applying a force field
- Crucial step and depends on the system
- **System preparation is a critical step to the success of the simulation**



1: System Preparation

► Parameters to generate the Amber topology:



Force_field: ff19SB

Water_type: TIP3P

Size Box (Angstroms):

Size_box:



ATTENTION: Give the concentration in Molar units, AMBER tleap will neutralize your system automatically:

Ions: NaCl

Concentration: " 0.15

[Show code](#)



1: System Preparation

▶ Show 3D structure

46s color: gray

show_sidechains:

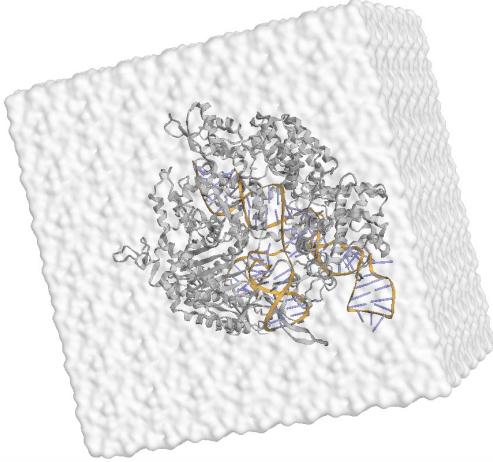
show_mainchains:

show_box:

box_opacity: —

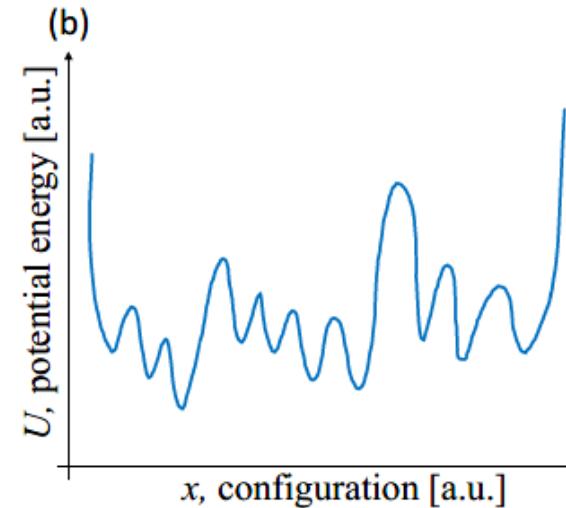
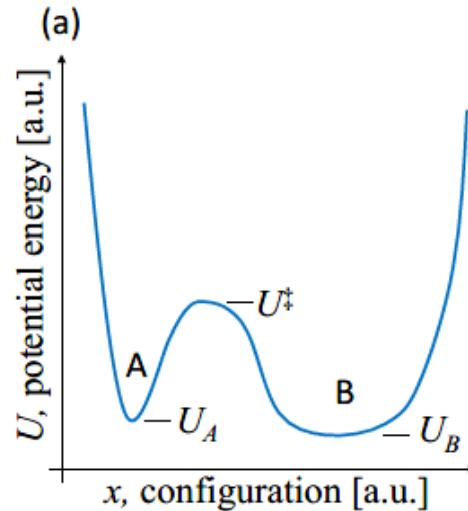
Show code

⟳



2: Energy Minimization

- MD minimization is done to obtain the local minimum energy structure of the molecule. This prevents the the system from “blowing up”
- Minimization results in a set of positions. The velocities will also need to be assigned.



3: Equilibration

- Goal: Bring the system to the target state point (energy, temperature, pressure) in a particular thermodynamic ensemble (NVE, NVT, NPT) [1].
- Thermostat will be needed to add or remove heat.
- System has reached equilibrium at the point of convergence. This is a sign that equilibration is complete.
- You want to sample in configurations in the state and ensemble that you want. This is the purpose of the equilibration step. To bring the system to the correct state and ensemble.



2: Energy Minimization + 3: Equilibration

Simulation time (in nanoseconds) and integration time (in femtoseconds):

Time: "0.2

Integration_timestep: 2

Temperature (in Kelvin) and Pressure (in bar)

Temperature: "298

Pressure: "1

Position restraints force constant (in kJ/mol):

Force_constant:



Frequency to write the trajectory file (in picoseconds):

Write_the_trajectory: 10

Frequency to write the log file (in picoseconds):

Write_the_log: 10

[Show code](#)

- ▶ Runs an Energy Minimization and Equilibration MD simulation (NPT ensemble)

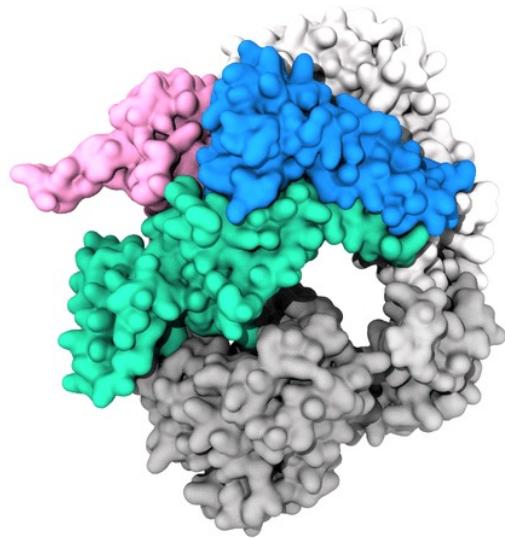
- Now, let's equilibrate our system!

[Show code](#)



4: Production Run

- During the production run, the data is collected.
- It is important to make sure that the system has converged before the production run begins
- Need to decide how often data is going to be stored
- Limitation: Disk Space



4: Production Run

Parameters for MD Production protocol:

Jobname: "Taki_prod"

Simulation time (in nanoseconds), number of strides (integers) and integration timestep (in femtoseconds):

Stride_Time: 0.2

Number_of_strides: 1

Integration_timestep: 2

Temperature (in Kelvin) and Pressure (in bar)

Temperature: 298

Pressure: 1

Frequency to write the trajectory file (in picoseconds):

Write_the_trajectory: 10

Frequency to write the log file (in picoseconds):

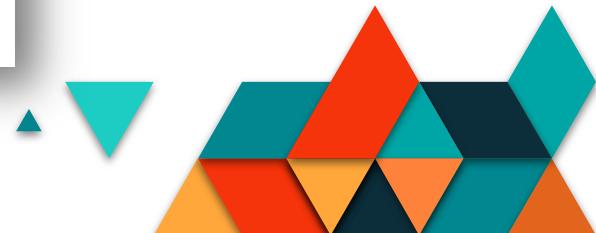
Write_the_log: 10

[Show code](#)

- Runs a Production MD simulation (NPT ensemble) after equilibration



[Show code](#)



4: Production Run

▶ Concatenate and align the trajectory

- ▶ **Important:** The Google Drive Path, Jobname, Number of strides, stride time and trajectory saved frequency should be the same you have been used to run your simulation in the previous steps.

Equilibrated_PDB: "1aki_equil.pdb

Jobname: "1aki_prod

Skip: 1

Output_format: dcd

first_stride: "1

Number_of_strides: "1

stride_time: "0.2

trajectory_saved_frequency: 10

Remove_waters: yes

[Show code](#)

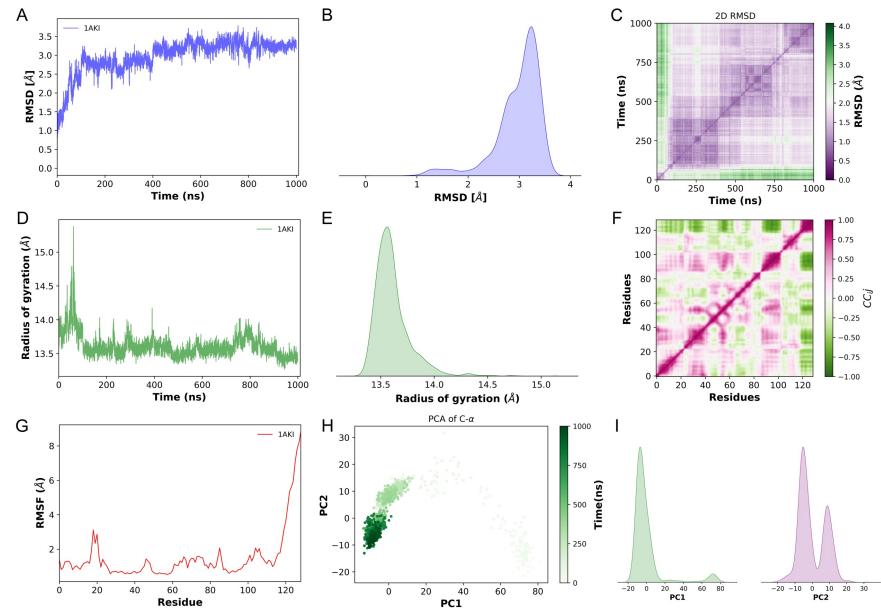


Common Analysis for MD Runs

Common Analysis Methods include:

- Interaction strength
- RMSD (a measure of structure stability)
- RMSF (a measure of structure stability)
- Radius of Gyration
- Principle Component Analysis
- Distance Analysis

Data can be analyzed using software-embedded packages, python libraries and many other codes.



Analysis for MD Runs

RMSD (Root Mean Square Deviation)

An example of RMSD data for a trajectory

Y-axis: RMSD value (unit: Angstrom)

X-axis: Time (ns)

RMSD calculations commonly done with the first frame as the frame of reference. This lets you know if the system has converged.

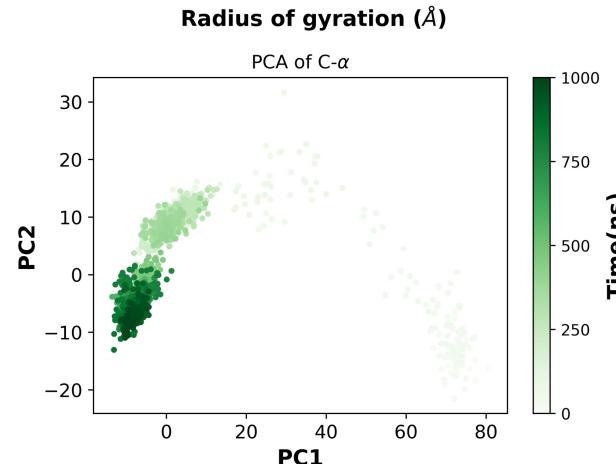
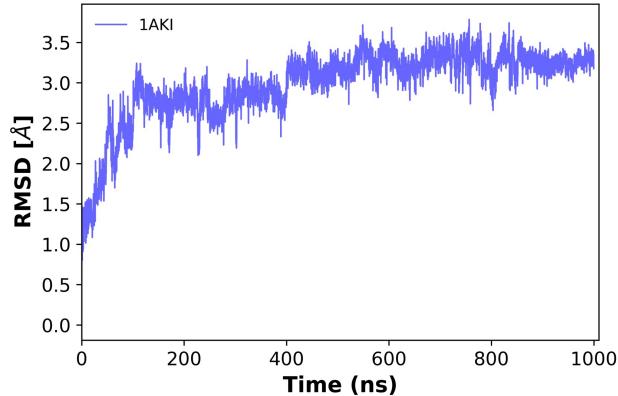
Other frames can be used as a reference as well.

Principle Component Analysis:

With the positional information provided by the trajectory.

PCA will allow you to determine which residues of your molecule have correlated movements.

This is done by calculating a coordinate covariance matrix, and then calculating principle components and coordinate projections.



Analysis for MD Runs

In the notebook you have several analysis, please, do all of them (Here the example is with RMSD):

▼ Analysis

Although visualizing your trajectory can be quite useful, sometimes you also want more quantitative data.

Analyses of MD trajectories vary a lot and we do not intend to cover it all here. However, one can make use of MDanalysis or PyTraj to easily analyze simulations.

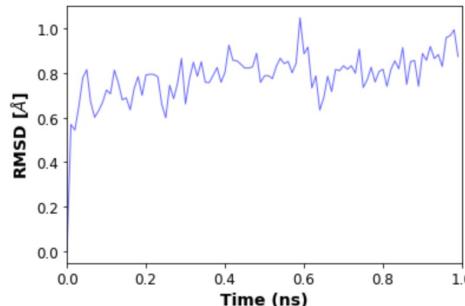
Below, you can find a few examples of code snippets that can help you to shed some light on your simulation behavior.

► Compute RMSD of protein's CA atoms

► Provide output file names below:

Output_name: "rmsd_ca"

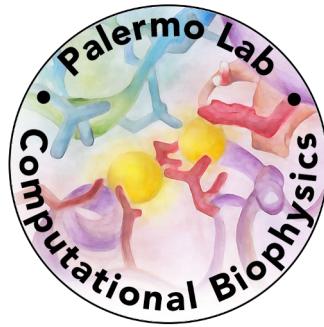
Show code



References

1. Braun, E. Gilmer, J. Mayes, H. Mobley, D. Monroe, J. Prasad, S. Zuckerman, D. (2018) Best Practices for Foundations in Molecular Simulations. *Living J. Comp. Mol. Sci.* Volume 1. Issue 1. 2019.
2. Ippoliti, E. What is Molecular Dynamics?. 2011
3. Alder, B. J.; Wainwright, T. E. (1959). "Studies in Molecular Dynamics. I. General Method". *J. Chem. Phys.* 31 (2): 459
4. L. Verlet, *Phys. Rev.* 159, p. 98 (1967).
5. M. Karplus, J. A. McCammon, *Scientific American* April 1986, p. 42 (1986).
6. Kuriyan, John. *The Molecules of Life: Physical and Chemical Principles* (p. 211).
7. Roe, D.R. *ambermd.org*. 2014. <http://ambermd.org/tutorials/analysis/tutorial1/index.htm>





Department of Bioengineering
University of California Riverside
223 Materials Science & Engineering
900 University Ave. | Riverside, CA 92521

<https://palermolab.com/>



website

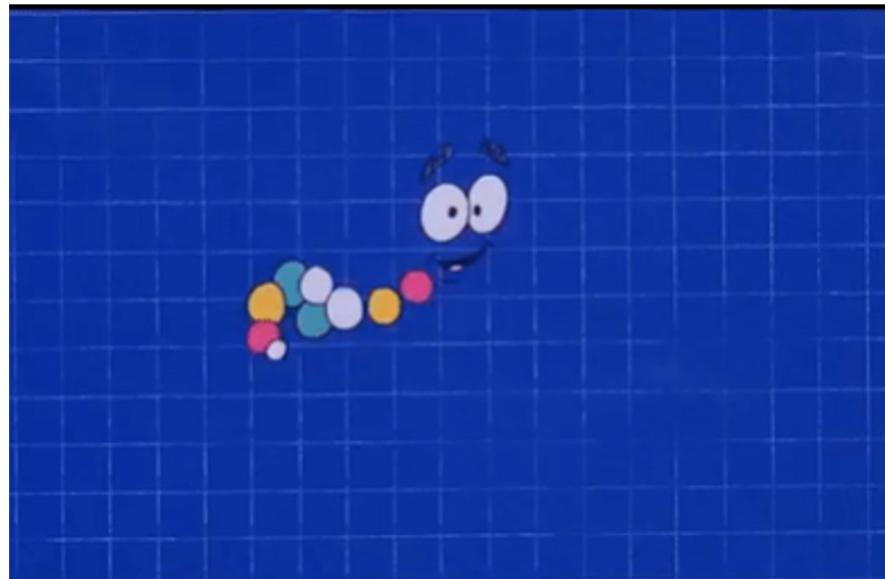


SCAN ME

pablitoarantes@gmail.com

pabloa@ucr.edu

<https://pablo-arantes.github.io>



twitter



SCAN ME



Obrigado (Thank you)! ☕

