800 ANNI 1222·2022

UNIVERSITÀ DEGLI STUDI DI PADOVA

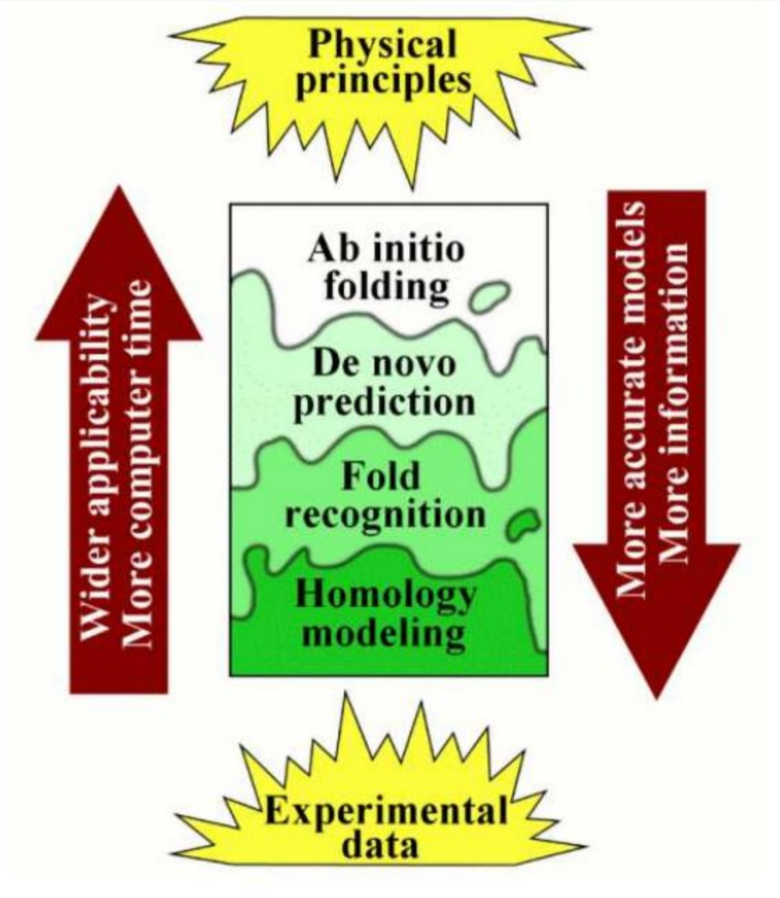
DIPARTIMENTO MATEMATICA

DATA SCIENCE
UNIVERSITY OF PADOVA

MOLECULAR DYNAMICS

Master of Science in Data Science

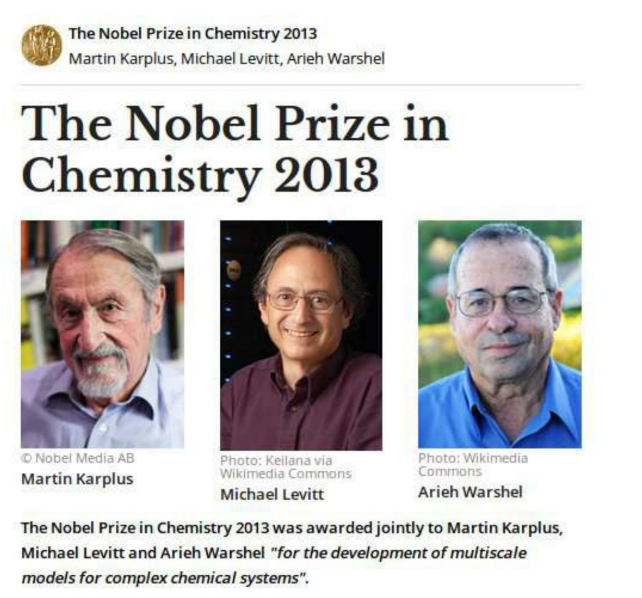
Damiano Piovesan



- **De novo prediction / Ab initio**
  - Secondary structure prediction; conformation of short fragments (Rosetta); molecular dynamics; Monte Carlo; quantum mechanics (unfeasible)
  - Tough computation
- **Fold recognition**
  - Try to fit with known folds
  - The fold space is not completely known (50% success)
- **Homology modelling**
  - Similar sequences have similar structures (+50% sequence identity)
  - 40% of genes are not homologous to known structures

Why study protein dynamics?

- "Flexibility" is a crucial aspect of protein function. An average protein structure alone cannot fully explain protein behavior
- Proteins are constantly exploring different configurations, which play a role in processes such as binding, catalysis, and allosteric regulation
- By analyzing the conformational changes between the active and inactive states of a protein, we can gain a better understanding of its function



The Nobel Prize in Chemistry 2013
Martin Karplus, Michael Levitt, Arieh Warshel

The Nobel Prize in Chemistry 2013

© Nobel Media AB
Martin Karplus

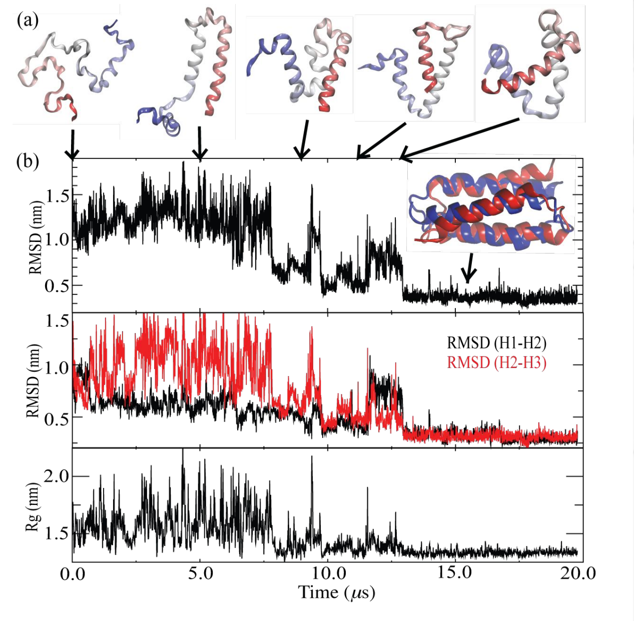
Photo: Keilana via Wikimedia Commons
Michael Levitt

Photo: Wikimedia Commons
Arieh Warshel

The Nobel Prize in Chemistry 2013 was awarded jointly to Martin Karplus, Michael Levitt and Arieh Warshel *"for the development of multiscale models for complex chemical systems"*.

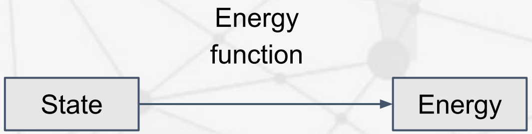
Fields of application

- Modelling of protein structures
- Conformational sampling and flexibility analysis for thermodynamics (free energy) studies
- Incorporation of solvent effects in NMR and crystallography experiments
- Study of protein-protein interactions and the effects of mutations on protein complexes



Energy functions

- A mathematical algorithm that inputs the **state of a system** and outputs the **energy**
- **System state** may include
  - Coordinates
  - Bonded structure
  - External elements (surface that repel, binding)
  - Empirical parameters (force fields, how energy works)
  - Charge, spin



- **Energy function** may include
  - Logic (if bonded…, if outside box…)
  - Arithmetic
  - Algebra (polynomial, exponential, log)
  - Calculus (derivatives, integrals)
  - Numerical procedures (advances, custom computer programs)
- **Energy function** can be
  - Simple or complex
  - Analytic or numerical
  - QM (law of physics)
  - MM (approximate) or other (empirical)
  - Ab initio or empirical
  - Atomistic or Coarse-grained

Molecular Mechanics (MM)

A set of models which use an **empirical** (lot of parameters), **algebraic** (formulas), **atomistic** (individual atoms) energy function for chemical systems

$$E_{total} = E_{bonded} + E_{non\text{-}bonded}$$

$$E_{bonded} = E_{bonds} + E_{angles} + E_{torsions}$$

$$E_{non\text{-}bonded} = E_{elst} + E_{vdw}$$

Force field models

- Atoms: 3D point particles
- Structure: bonds, angles, torsions

Molecular Mechanics (MM)

Force fields vary in

- Energy function form
- Empirical parameters
- Simulation targets

Limitations

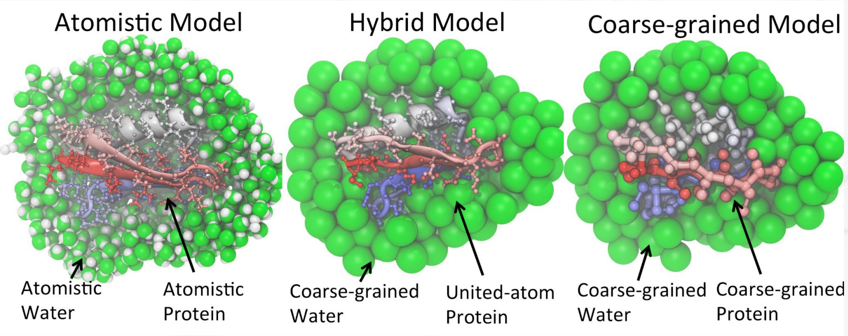
- Up to 1M of atoms
- Up to 10^{15} configurations
- Accuracy of energy
- Scope of parameters

Examples

- **AMBER**, assisted model building energy refinement (originally for DNA)
- **CHARM**, chemistry at harvard MM (Martin Karplus nobel 2013)
- **OPLS**, optimized potential for liquid simulations (small molecules)
- **GROMOS**, used by Gromacs software
- **MMFT**, Merk Molecular force field (pharmaceutical)

Types of force fields

- **All-atom** force fields have parameters for every atom species, including mass, charge, and volume
- **United-atom** force fields treat hydrogen and carbon atoms as methyl and methylene groups, respectively, reducing the number of centers of interaction
- **Coarse-grained** force fields simplify the system by representing multiple atoms as spheres, allowing for simulations with longer timescales, often in the microseconds or longer range, while sacrificing some detail. This makes them computationally efficient.



Force field parameters

An arbitrary constant whose value characterize an element of a system

$$E_{total} = \sum_{bonds} K_b (r - r_{eq})^2 + \ldots$$

Units

E → [kcal/mol]   r → [Å]

R_{eq} → [Å] K_b → [kcal/mol x Å^2]

Example

O-H in H_2O in AMBER95

R_{eq} = 0.960 Å k_b = 553.0 kcal/mol x Å^2

Sources

- Structures → crystal, electron density
- Spectra (NRM) → IR (vibration freq. of bonds)
- Simulation → PES (potential energy surface, quantum mechanics)

Desired properties

- General
- Transferable
- Accurate

Bond stretch terms

$$E_{bonds} = \sum_{bonds} K_b(r - r_{eq})^2$$

Parameters

r_{eq} → equilibrium bond length

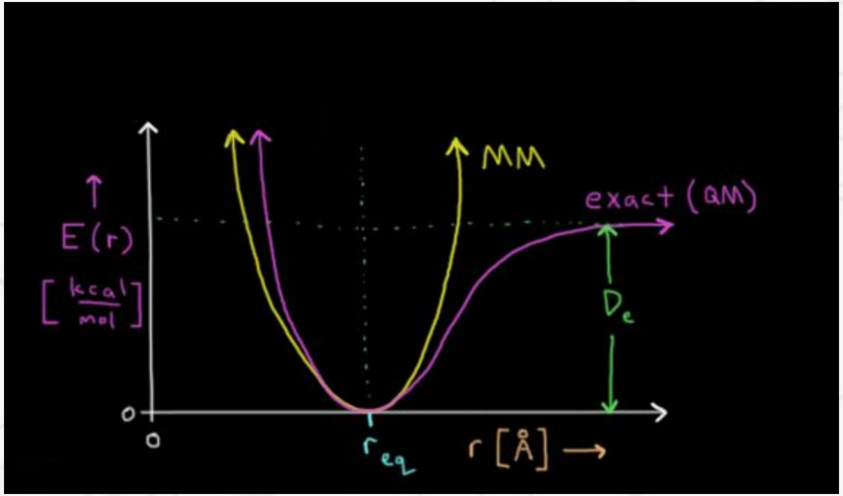
K_b → bond spring constant

Variables

r → bond length

D_e → dissociation energy

Typical $K_b \approx$ 300-600 kcal/mol x Å^2



Angle bend terms

$$E_{angles} = \sum_{angles} K_a(\theta - \theta_{eq})^2$$

Parameters

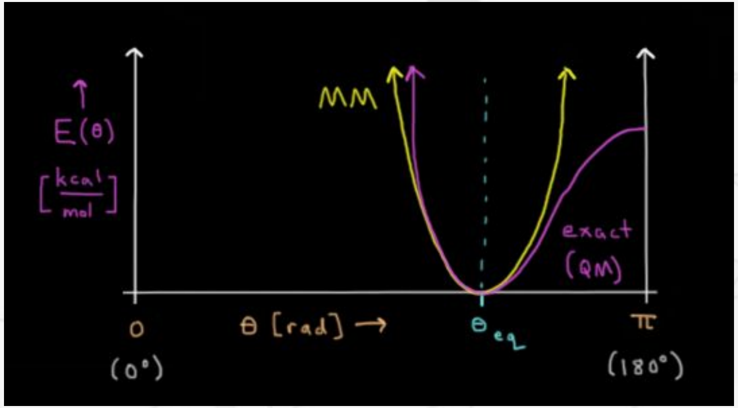
θ_{eq} → equilibrium bond angle [rad]

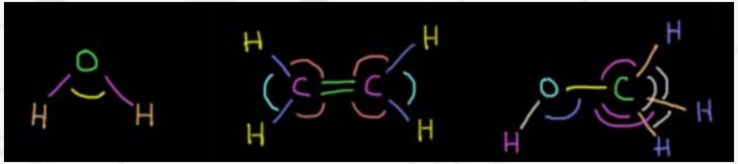
θ_a → angle spring constant

Variables

θ → bond angle [rad], $0 \leq \theta \leq \pi$

Typical K_a ≈ 100-150 kcal/mol x rad^2





Van der Waals (non bonded) terms

$$E_{vdw} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \varepsilon_{ij} \left[\left(\frac{r_0}{r_{ij}}\right)^{12} - 2\left(\frac{r_0}{r_{ij}}\right)^{6} \right]$$

Parameters

- ε_{ij} → interaction strength [kcal/mol]
- r_0 → van der Waals radius (Å)

Variables

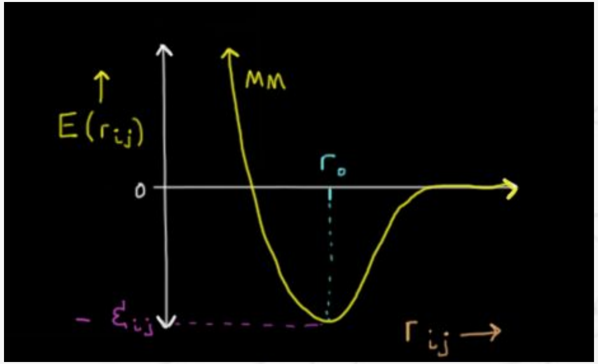
- r_{ij} → interatomic distance (Å)

Short range ($r_{ij} < r_0$)

- Exchange / steric repulsion
- "^12" term dominates

Long range ($r_{ij} > r_0$)

- London dispersion attraction
- "^6" term decay more slowly



How to define parameters for all possible pairs?

"Combining rules" → ε_i, ε_j atomic parameters

$\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}$ "geometric mean"

$r_0 = ½(r_i + r_j)$ "arithmetic mean"

Electrostatics terms

$$E_{elst} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

Parameters

q_i , q_j → atomic partial charges (e)

ε_0 permittivity of free space (constant)

Variables

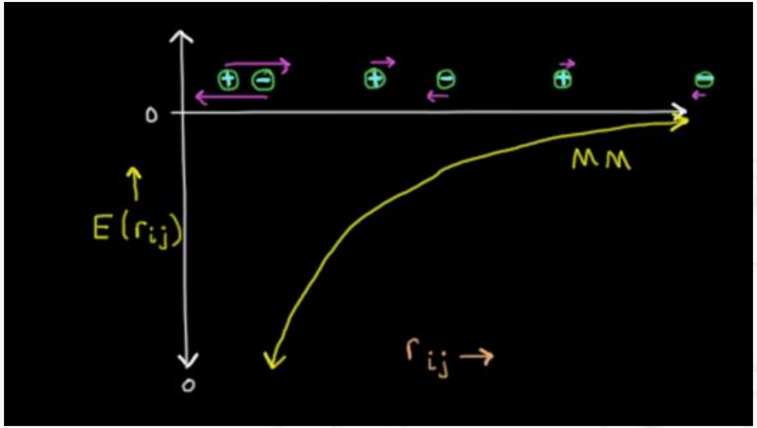
r_{ij} → interatomic distance (Å)

Same charge → repulsion

Opposite charge → attraction

Charge of the electron (e), 1.6×10^{-19} coulomb

Q_{tot} usually is zero



vdw $\propto \left(\frac{1}{r}\right)^6$ → quick decay

elst $\propto \left(\frac{1}{r}\right)^1$ → slow decay

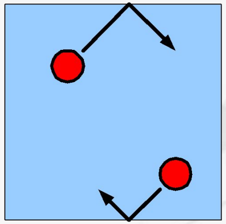
Elst is rate limiting step of MM

- O(N) bonded terms
- $O(N^2)$ non-bonded terms

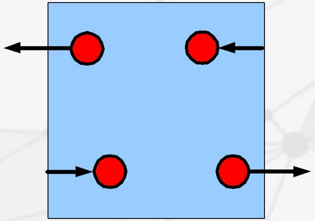
Boundary conditions

The goal is to simulate an infinite system while using a finite simulation box

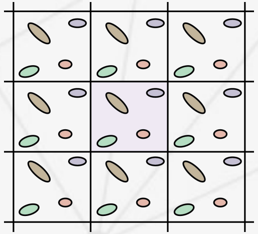
- Particles experience forces as if they were in a bulk fluid
- If a particle exits the simulation box, it is replaced by an image particle that enters from the opposite side. This maintains the continuity of the system and prevents the introduction of artificial boundary effects



Specular boundary condition



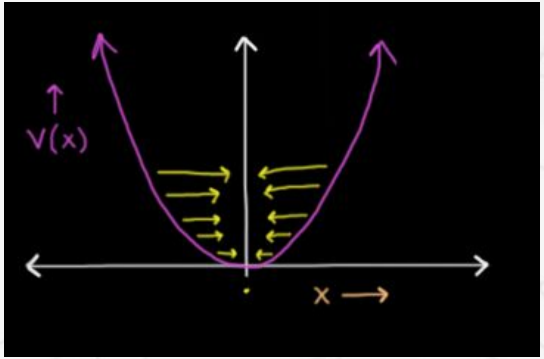
Periodic boundary condition



Energy gradient

$V(x)$ → potential energy (example: $\frac{1}{2} kx^2$)

$F(x) = -\frac{dV(x)}{dx}$ → force (example: $-kx$)



1D		**3D**
V(x)	→	V(x,y,z)
N atoms	→	3N coordinates

Gradient

$$-\frac{dV(x)}{dx}$$

$$\underbrace{-\left(\frac{\partial V}{\partial x}\hat{x} + \frac{\partial V}{\partial y}\hat{y} + \frac{\partial V}{\partial z}\hat{z}\right)}_{-\nabla V(x,y,z)}$$

Notation

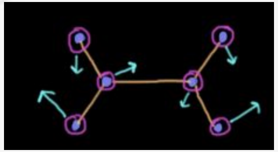
$$V = V(x_1, y_1, z_1, x_2, y_2, z_2, \ldots, x_N, y_N, z_N)$$

$$= V(\bar{r}_1, \bar{r}_1, \ldots, \bar{r}_N) \qquad \bar{r}_1 = (x_i, y_i, z_i)$$

$$= V(\bar{r}^{3N})$$

$V(\bar{r}^{3N})$ → energy

$-\nabla V(\bar{r}^{3N})$ → force

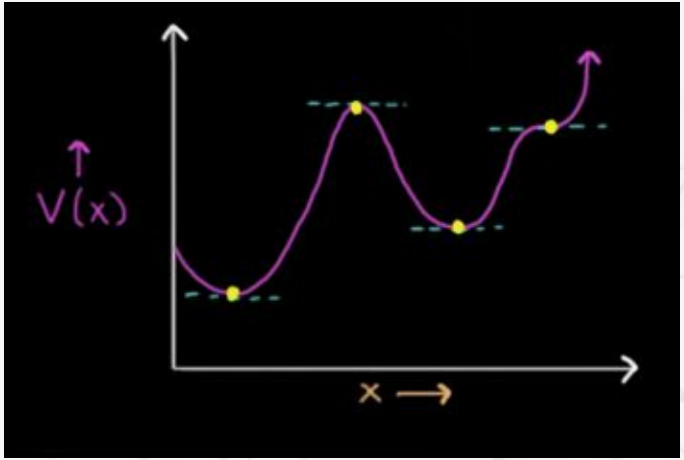


DIPARTIMENTO
MATEMATICA

Energy minimization

$$\frac{dV(x)}{dx} = 0$$ → “stationary point”

$$\frac{d^2V(x)}{dx^2} \begin{cases} > 0 \text{ local minimum} \\ < 0 \text{ local maximum} \\ = 0 \text{ unknown (“saddle” point)} \end{cases}$$



1D → **3D**

$$\frac{dV(x)}{dx} = 0 \qquad \left|\nabla V(\bar{r}^{3N})\right| = 0$$

$$\frac{d^2V(x)}{dx^2} > 0 \qquad \{\lambda_H\} \geq 0,\ H_{ij} = \frac{\partial^2 V(\bar{r}^{3N})}{\partial q_i \partial q_j}$$

H → 3N x 3N real, symmetric “Hessian matrix”
Local minima → All eigenvalues are ≥ 0

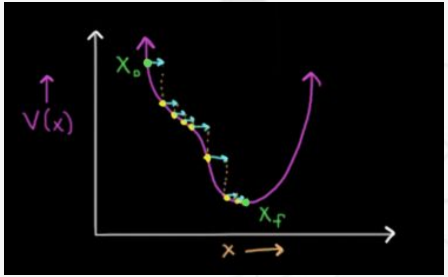
- Lowest $V(\bar{r}^{3N})$ among local minima → global minimum
- “Geometry Optimization” → algorithm to find the coordinates $\{\bar{r}^{3N}\}$ with the minimum energy

Steepest descent

Method for obtaining local stationary points of a multi-dimensional function

Simple method for geometry optimization

1. Guess X_0 “initial geometry”
2. Compute $V(x_i)$ and $\nabla V(x_i)$, energy and gradient
3. Update $x_{i+1} = x_i - \gamma_i \nabla V(x_i)$, displacement
4. Check for convergence
    a. If yes: done $x_f = x_i$
    b. If no: return to step 2



There are various methods for choosing γ_n

Convergence checks

$\lVert \nabla V(x_i) \rVert \approx 0$	gradient
$\lvert V(x_i) - V(x_{i-1}) \rvert \approx 0$	energy
$\lVert x_i - x_{i-1} \rVert \approx 0$	displacement

Molecular dynamics (MD)

Calculate system properties by propagating through **time** and averaging

- Classical Mechanics → Newton’s Laws (F = ma)
- MM → conservative (total energy is conserved, kinetic + potential energy), time-independent

$$F = ma = m\frac{d^2(x)}{dt^2} = \frac{d}{dt}\left(m\frac{d(x)}{dt}\right) = \frac{d}{dt}(mv) = \frac{dp}{dt}$$

$$F = -\frac{dV(x)}{dx}$$

$$\frac{\partial p_{x_i}}{\partial t} = -\frac{\partial V}{\partial x_i}$$

$$p = mv = m\frac{dx(t)}{dt}$$

f = force
m = mass
a = acceleration
x = position / coordinates
p = momentum

Solve for all 3N coordinates

Can’t solve exactly → approximate numerically

Molecular dynamics

1. Assign initial x_0, p_0, $t = 0$
2. Compute $V(x_i)$ and $\nabla V(x_i)$
3. Update $x_{i+1} = x_i + 1/m\, p_i \Delta t$
4. Update $p_{i+1} = p_i - \nabla V(x_i) \Delta t$
5. Repeat (steps 2-4) until $t \geq t_f$
6. Compute desidered properties

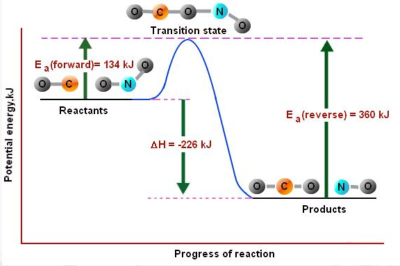
$$\left.\begin{array}{l} 3. \\ 4. \end{array}\right\} t = t + \Delta t$$

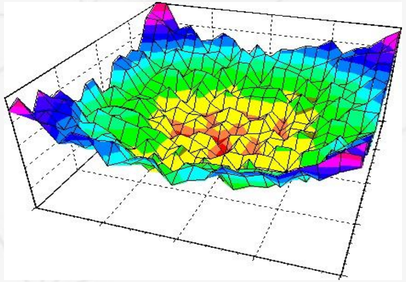
Δt → timestep ($1\text{fs} = 10^{-15}\text{s}$)

t = time
x = position / coordinates
p = momentum (mv)
m = mass
V = potential energy
∇V = energy gradient

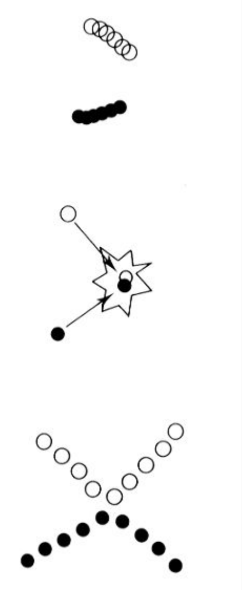
Minimization Vs Molecular Dynamics

- All molecules can exist with relative atom placements that are energetically unfavorable, i.e. it requires input of energy to put a molecule into a particular conformation
- **Energy minimization** finds the closest minimal energy conformation, but may only identify **local minima** and not the global minimum
- The goal of **molecular dynamics** is not to reduce energy to the closest minimum energy conformation but to more or less **conserve energy** while allowing atoms to move about (while constrained by the forcefield)
- If allowed to proceed long enough, MD can sample different molecular conformations, especially if the amount of starting energy is high (e.g. "high temperature")



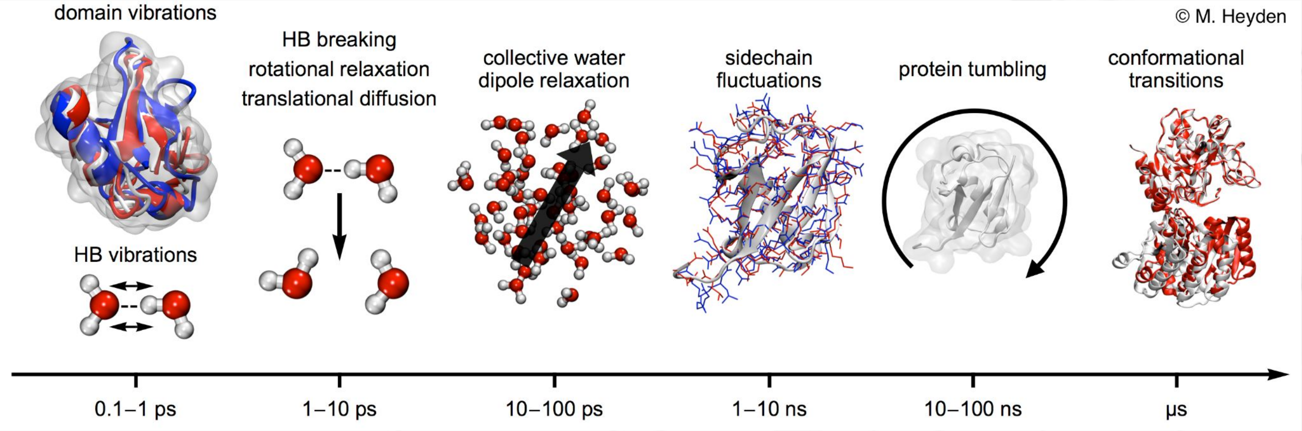


The right simulation time



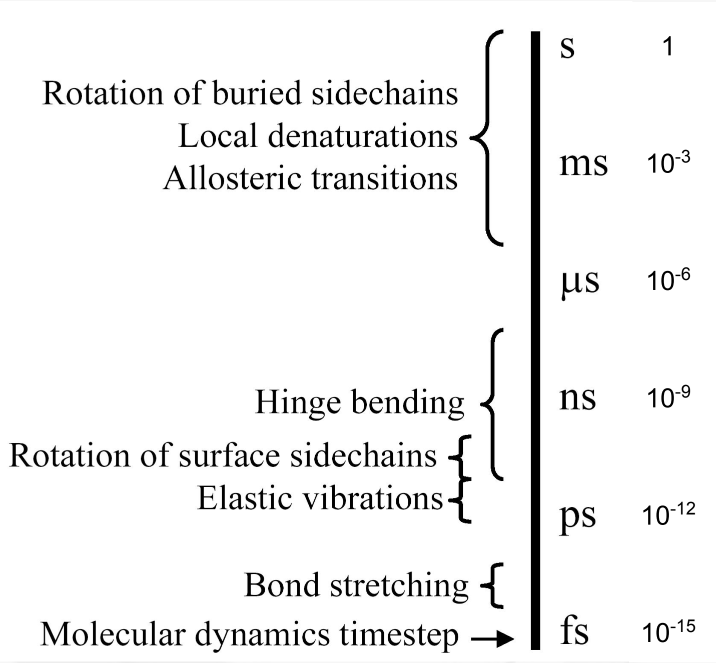
1. Too short – no variations are appreciated
2. Too long – errors derived from approximations
3. Fair compromise – acceptable errors, maximum speed

Time scales



10^{-13} - 10^{-12} 10^{-12} - 10^{-11} 10^{-11} - 10^{-10} 10^{-9} - 10^{-8} 10^{-8} - 10^{-7} 10^{-6}

Time scales



Final T → depends on goal

- Vibrations → 10 - 1,000 fs
- Conformation changes → ps - ns
- Diffusion / interaction → ns - µs
- Protein folding → µs - s

Molecular Dynamics & Biology

- What can we simulate?
- What kind of information do we get?
- How can we use them to solve a biological problem?

References & Links

TMP Chem (Trent Parker’s YouTube channel)

https://www.youtube.com/user/TMPChem

Computational Chemistry

- 2.1, 2.2, 2.3
- 2.5, 2.6
- 2.9, 2.10
- 3.1, 3.2, 3.3
- 3.8

800 ANNI 1222·2022

UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO MATEMATICA

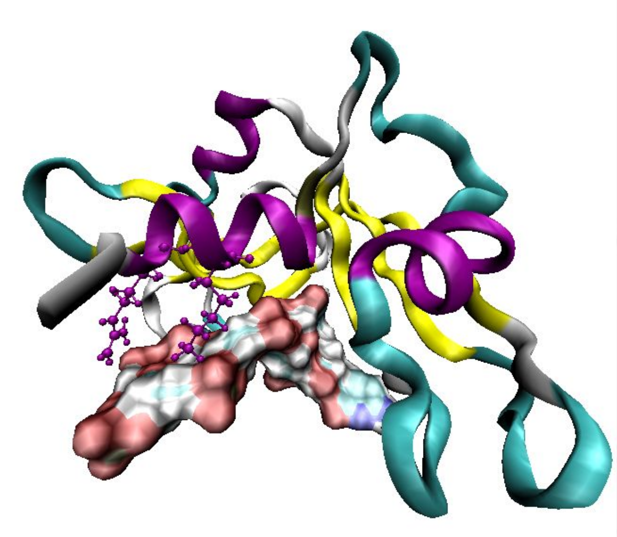
DATA SCIENCE
UNIVERSITY OF PADOVA

Biological Example:
Human pancreatic ribonuclease

Master of Science in Data Science

Damiano Piovesan

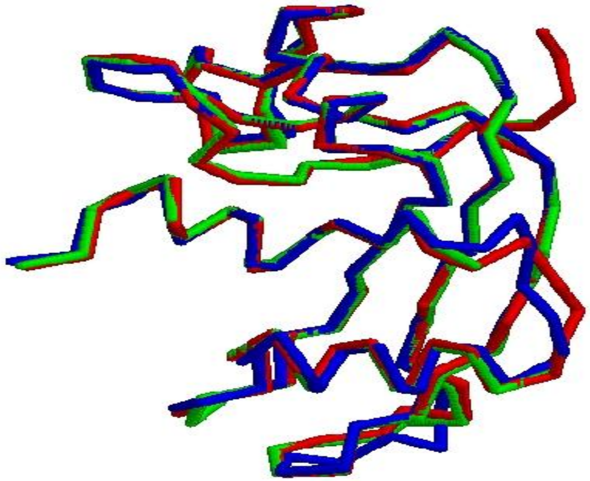
Human pancreatic ribonuclease



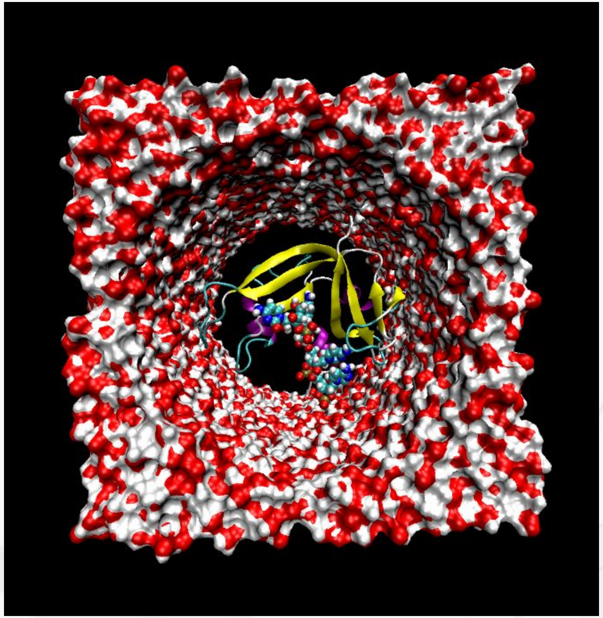
- Comparative study between the human pancreatic enzyme and Ribonuclease A
- MD to validate the reliability of a model

Simulation setup

- Initial structures
  - RNAse A bovine, 1RCN
  - The mutant HP-RNase, 1DZA
- Tested bound with 4 DNA nucleotides (ATAA)
- The structures are all very similar

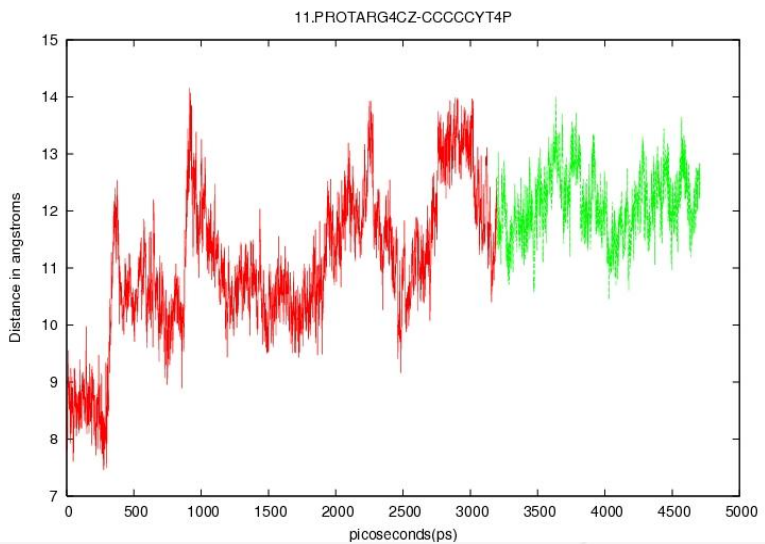


Simulation parameters

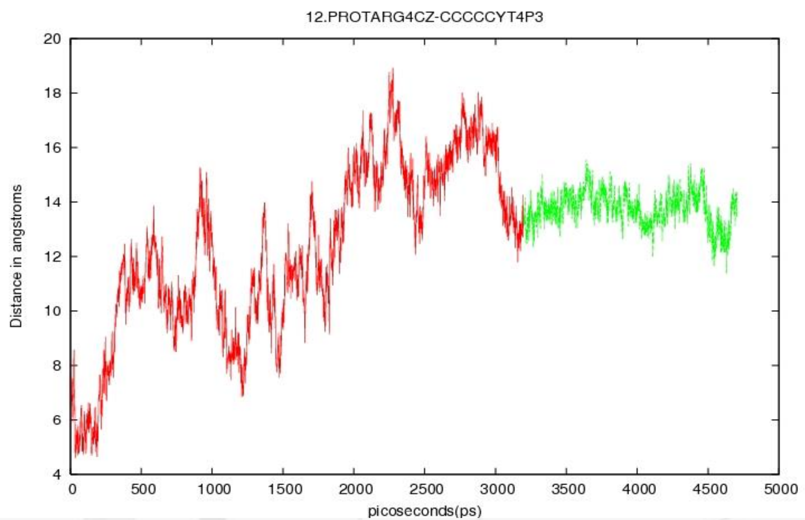


- Waterbox 65 Å (~30,000 atoms)
- Charmm force field, no extra parameters. NAMD/Vmd software
- PBC (periodic boundary conditions) with Particle Mesh Ewald for electrostatics
- 2 fs timestep, shake, NPT (constant-temperature/pressure) ensemble
- Progressive heating and no restriction for almost 1ns
- Simulation ~500ns

Analysis of a MD simulation



HP-RNase



bovine RNAse A

GROMACS

Master of Science in Data Science

Damiano Piovesan

Gromacs

- GROningen MAchine for Chemical Simulations
- Open source implementations of Gromos integrator
    - www.gromacs.org

FAST. FLEXIBLE. FREE.
GROMACS