

# Computational Chemistry - MD Simulations

P. Ojeda-May

[pedro.ojeda-may@umu.se](mailto:pedro.ojeda-may@umu.se)

High Performance Computing Center (HPC2N),  
Umeå University,



901 87, Sweden.

# Table of contents

# Particle Dynamics

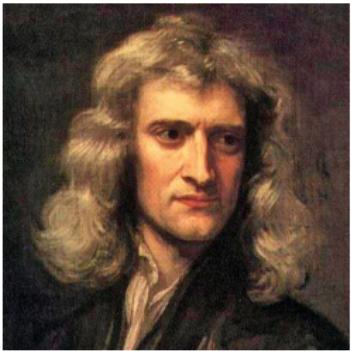
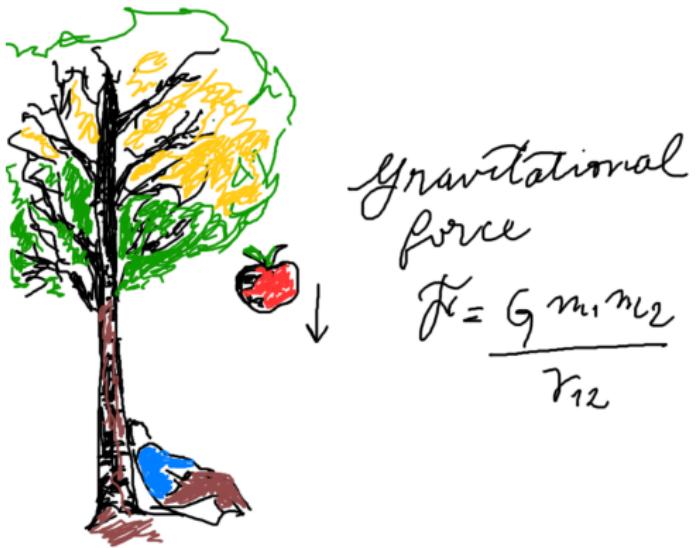


Figure: Isaac Newton  
(1642-1726). (Source: Wikipedia)



# Newton's equation

$$\mathbf{F} = -\nabla U \quad \text{Newton's Law(1687)} \quad (1)$$

solution of this equation requires the knowledge of an array of particles' positions and velocities

$$\mathbf{X} = (x_1^1, x_2^1, x_3^1, x_1^2, x_2^2, x_3^2 \dots x_1^N, x_2^N, x_3^N) \quad (2)$$

$$\mathbf{V} = (v_1^1, v_2^1, v_3^1, v_1^2, v_2^2, v_3^2 \dots v_1^N, v_2^N, v_3^N) \quad (3)$$

# Protein Conformational Space

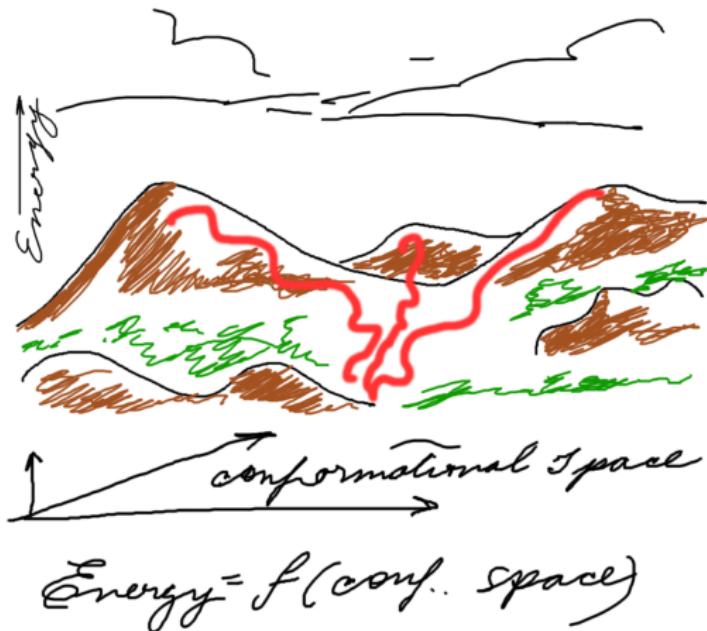


Figure: Energy landscape.

# Early MD simulations

## Phase Transition for a Hard Sphere System

PHYSICAL REVIEW

VOLUME 136, NUMBER 2A

19 C

B. J. ALDER AND T. E. WAINWRIGHT

University of California Radiation Laboratory, Livermore, California

(Received August 12, 1957)

## Correlations in the Motion of Atoms in Liquid Argon\*

A. RAHMAN

Argonne National Laboratory, Argonne, Illinois

(Received 6 May 1964)

## Computer simulation of protein folding

Michael Levitt\* & Arieh Warshel\*

Department of Chemical Physics, Weizmann Institute of Science, Rehovot, Israel

*A new and very simple representation of protein conformations has been used together with energy minimisation and thermalisation to simulate protein folding. Under certain conditions, the method succeeds in 'renaturing' bovine pancreatic trypsin inhibitor from an open-chain conformation into a folded conformation close to that of the native molecule.*

protein, in this case myoglobin, was based on the packing of cylinders supposed to represent  $\alpha$  helices<sup>4</sup>. The method was not implemented on a computer and cannot be applied more generally to other proteins not built entirely from helices.

Here we tackle the problem differently. First, we simplify the representation of a protein by averaging over the fine details. This is done both to make the calculations much more efficient and also to avoid having to distinguish between many

Figure: Nature, 253 (1975).

## The Nobel Prize in Chemistry 2013



Photo: A. Mahmoud  
**Martin Karplus**  
Prize share: 1/3



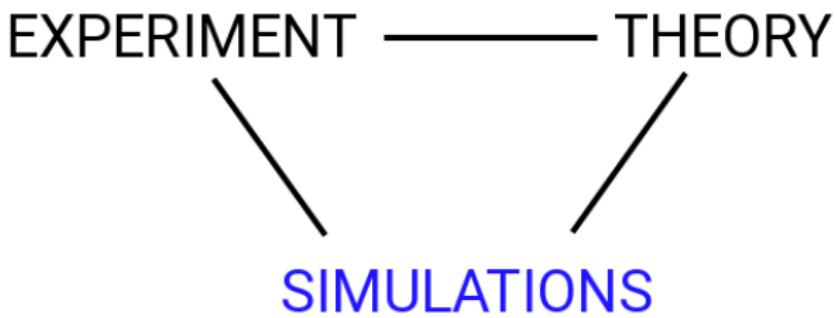
Photo: A. Mahmoud  
**Michael Levitt**  
Prize share: 1/3



Photo: A. Mahmoud  
**Arieh Warshel**  
Prize share: 1/3

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"*.

**Figure:** Credits: <http://www.nobelprize.org>.



# Application of MD

## Proteins

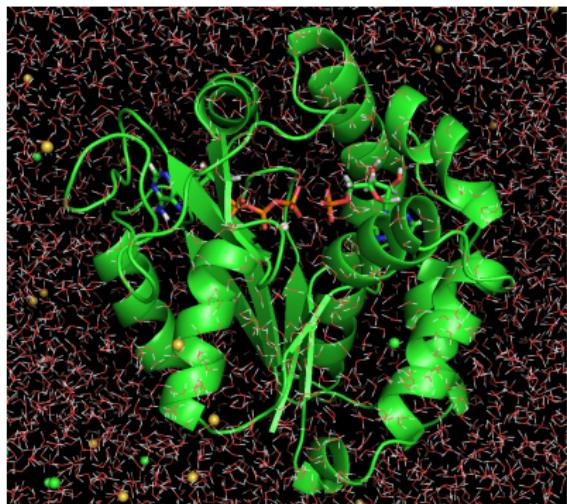


Figure: AdK enzyme in water.

## Clays

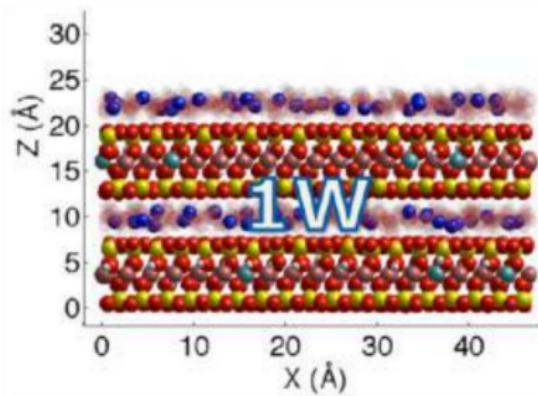


Figure: Clay [JPC C, 118, 1001 (2014)].

# Application of MD

Food Biophysics (2009) 4:340–346  
DOI 10.1007/s11483-009-9132-9

ORIGINAL ARTICLE

## Estimation of Water Diffusion Coefficients in Freeze-Concentrated Matrices of Sugar Solutions Using Molecular Dynamics: Correlation Between Estimated Diffusion Coefficients and Measured Ice-Crystal Recrystallization Rates

Tomoaki Hagiwara · Takaharu Sakiyama ·  
Hisahiko Watanabe

**Figure:** Ice cream research.

# Application of MD



Molecular dynamics study of interfacial mechanical behavior between asphalt binder and mineral aggregate

Guangji Xu, Hao Wang\*

Department of Civil and Environmental Engineering, Rutgers University, Piscataway, NJ 08854, USA

Figure: Asphalt research.

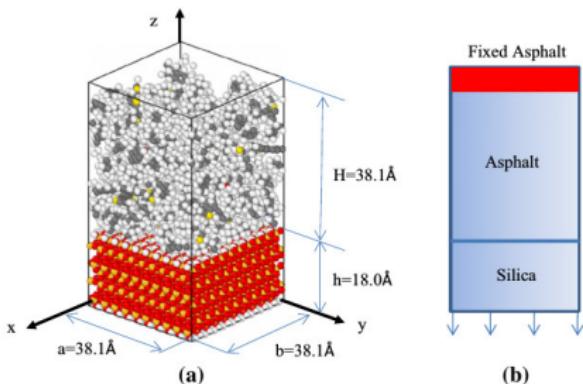


Fig. 3. Representative model for the MD simulations and schematic of tensile simulation.

Figure: Asphalt [Const. Build. Mat., 121, 246 (2016)].

# Force fields

$$U = \sum_{\text{bonds}} \frac{1}{2} k_{\text{bonds}} (r - r_0)^2 + \sum_{\text{angles}} \frac{1}{2} k_{\text{angle}} (\theta - \theta_0)^2 \\ + \sum_{\text{torsions}} \sum_j V_j (1 + \cos j\phi) \\ + \sum_{\text{Coulomb}} \frac{q_i q_j}{r_{ij}} + \sum_{\text{VdW}} \left\{ 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \right\} \quad (4)$$

# Force fields

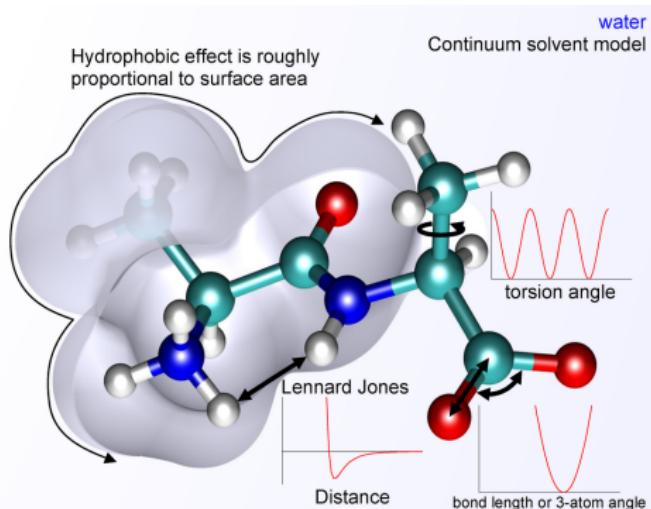


Figure: Energy terms. Credits: [https://en.wikipedia.org/wiki/Force\\_field\\_\(chemistry\)](https://en.wikipedia.org/wiki/Force_field_(chemistry))

# Force fields

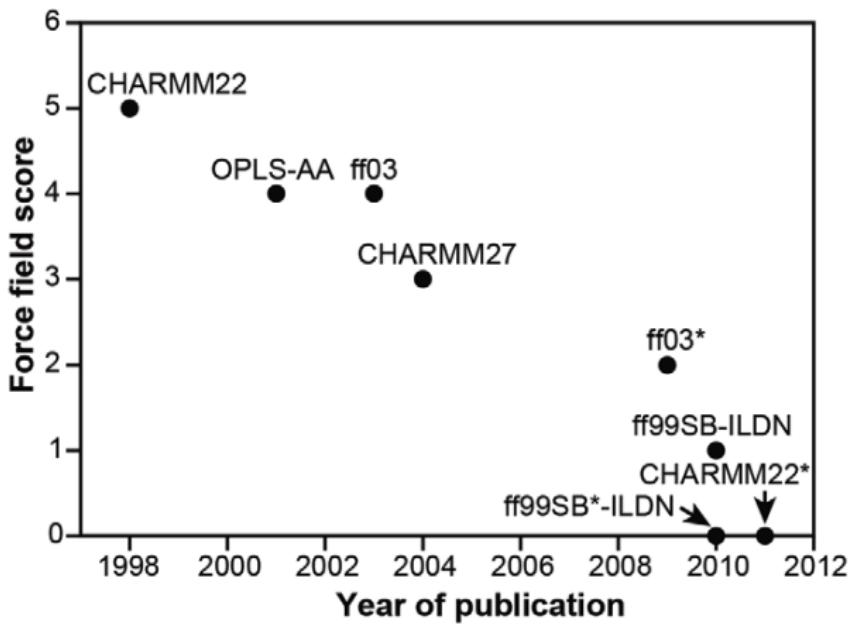


Figure: FF for proteins comparison, PLoS ONE, 7, e32131, (2012).

# Force fields: Energy surface

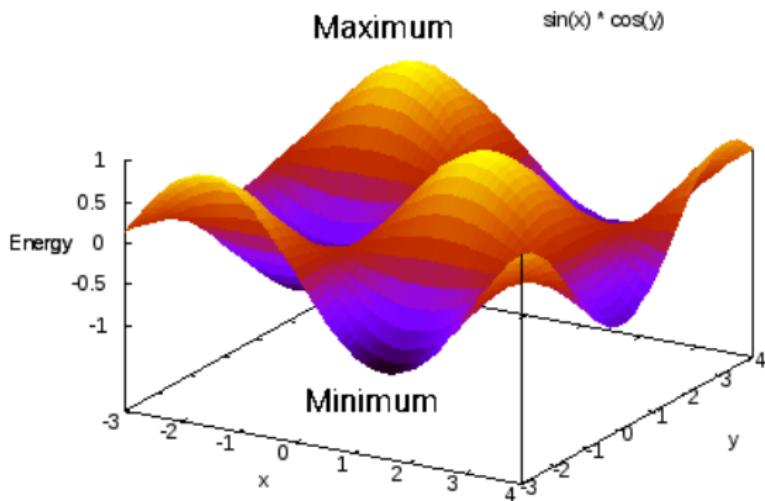


Figure: Energy surface described by  $U = \sin(x) * \cos(y)$

# Force fields

- Proteins and Hydrocarbons: GROMOS, OPLS-AA, AMBER, CHARMM.
- Clays: CLAYFF
- Coarse-graining: MARTINI

If the parameters of your compound are not part of the force field you need to use QM approaches.

# Water models

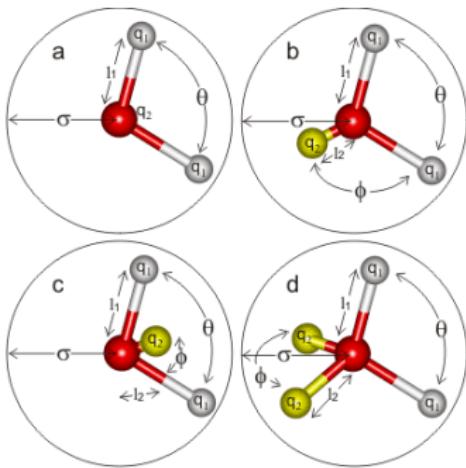


Figure: 3-5 sites water models. Credits:

[http://www1.lsbu.ac.uk/water/water\\_models.html](http://www1.lsbu.ac.uk/water/water_models.html)

# Protein systems

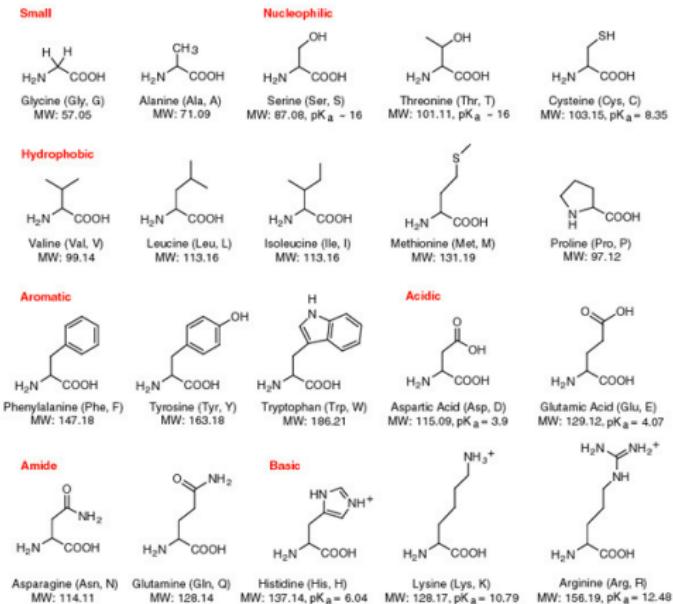


Figure: 20 natural amino acids. Credits: [goo.gl/YrYvvw](http://goo.gl/YrYvvw)

# Protein systems

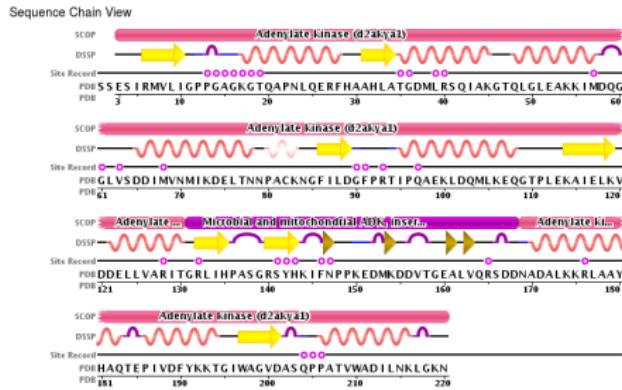


Figure: PDB information of AdK.

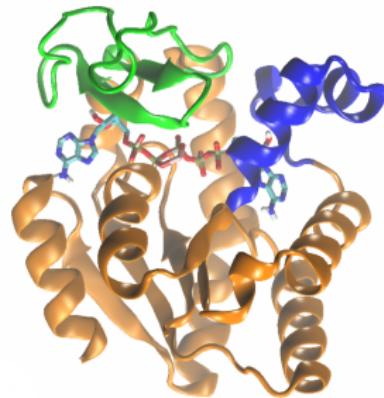


Figure: Structure of yeast AdK.

# Periodic boundary conditions (PBC)

The systems we can study with MD simulations are tiny compared to real experimental setups ( $10^{23}$  particles).

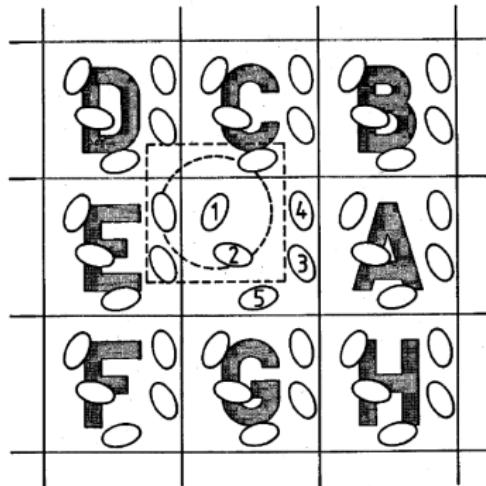


Figure: PBCs and minimum image convention [Allen & Tildesley, Comp. Sim. of Liquids]

# Periodic boundary conditions (PBC)

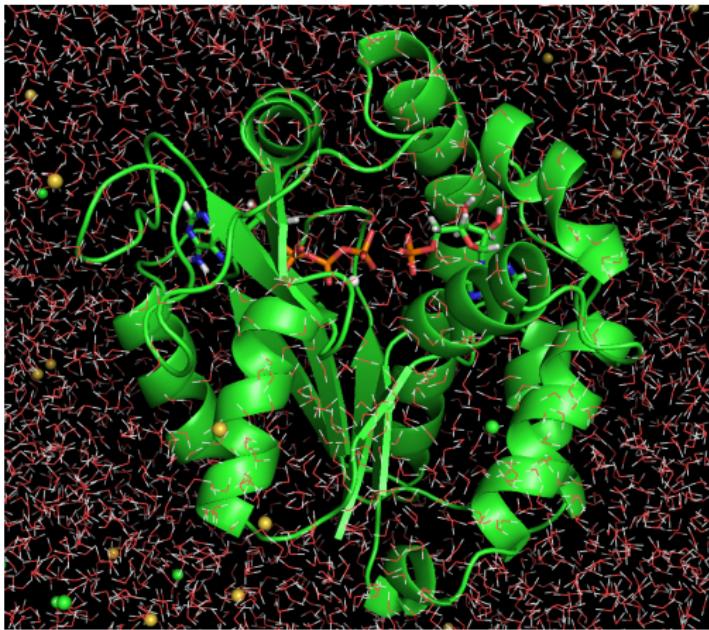


Figure: AdK enzyme in water.

# Electrostatic interactions: Ewald method

The electrostatic energy for a periodic system can be written as<sup>1</sup>,

$$E = \frac{1}{2} \sum_{m \in \mathbb{Z}^3}^{\infty} \sum_{i,j=1}^N ' \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{m}L|} \quad (5)$$

where  $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ ,  $\mathbf{m}$  refers to the periodic images. Primed summation means  $i = j$  interaction is excluded for  $\mathbf{m} = 0$ .  $q_x$  is the partial charge on atom  $x$ .

---

<sup>1</sup>Adv. Polym. Sci., 185, 59 (2005)

# Integration of Newton's equation

We now know the force field and we know the law of motion:

$$\mathbf{F} = m\mathbf{a} = -\nabla U \quad \text{Newton's Law} \quad (6)$$

we need to integrate this equation, here we use the leap-frog scheme [Hockney, 1970] ,

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \delta t \mathbf{v}(t + \frac{1}{2}\delta t) \quad (7)$$

$$\mathbf{v}(t + \frac{1}{2}\delta t) = \mathbf{v}(t - \frac{1}{2}\delta t) + \delta t \mathbf{a} \quad (8)$$

velocities are updated according to,

$$\mathbf{v}(t) = \frac{1}{2} \left( \mathbf{v}(t + \frac{1}{2}\delta t) + \mathbf{v}(t - \frac{1}{2}\delta t) \right) \quad (9)$$

# Ergodicity

$$\begin{aligned}\mathcal{A}_{obs} &= \langle \mathcal{A} \rangle_{\text{time}} \\ &= \langle \mathcal{A}(\Gamma(t)) \rangle_{\text{time}} \\ &= \lim_{t_{\text{obs}} \rightarrow \infty} \int_0^{t_{\text{obs}}} \mathcal{A}(\Gamma(t)) dt\end{aligned}\quad (10)$$



Figure: Coffee cup.

# Protein Conformational Space

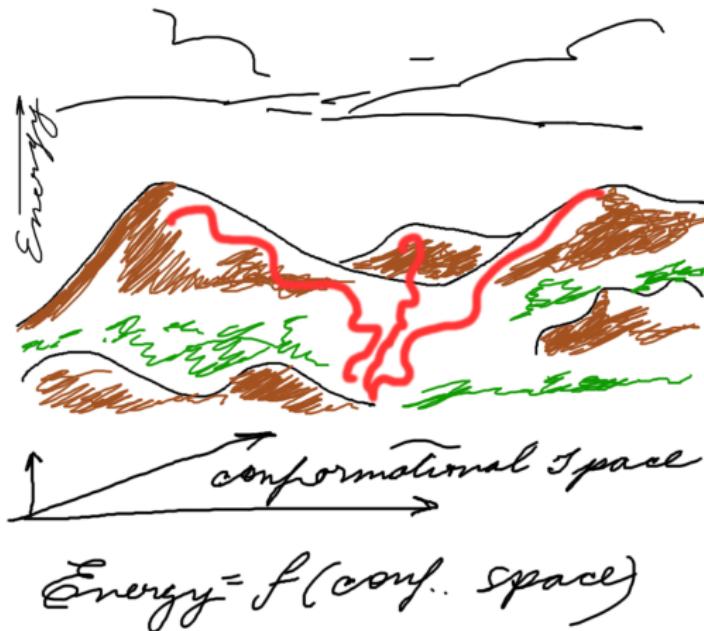


Figure: Energy landscape.

# Statistical ensembles

- Microcanonical ensemble (NVE) partition function is [Allen & Tildesley, Comp. Sim. of Liquids],

$$Q_{NVE} = \frac{1}{N!} \frac{1}{h^{3N}} \int d\mathbf{r} d\mathbf{p} \delta(\mathcal{H}(\mathbf{r}, \mathbf{p}) - E) \quad (11)$$

The thermodynamic potential is the negative of the entropy  
 $-S/k_B = -\ln Q_{NVE}$

- In the case of the Canonical ensemble (NVT) the partition function is,

$$Q_{NVT} = \frac{1}{N!} \frac{1}{h^{3N}} \int d\mathbf{r} d\mathbf{p} \exp(-\mathcal{H}(\mathbf{r}, \mathbf{p})/k_B T) \quad (12)$$

with thermodynamic potential  $A/k_B T = -\ln Q_{NVT}$ .

# Statistical ensembles

- Isothermal-isobaric ensemble (NPT) partition function is,

$$Q_{NPT} = \frac{1}{N!} \frac{1}{h^{3N}} \frac{1}{V_0} \int dV \int d\mathbf{r} d\mathbf{p} \exp(-(\mathcal{H}(\mathbf{r}, \mathbf{p}) + PV)/k_B T) \quad (13)$$

the corresponding thermodynamic potential is

$$G/k_B = -\ln Q_{NPT}$$

- Grand-canonical ensemble ( $\mu VT$ ) partition function is,

$$Q_{\mu VT} = \sum_N \frac{1}{N!} \frac{1}{h^{3N}} \exp(\mu N/k_B T) \int d\mathbf{r} d\mathbf{p} \exp(-\mathcal{H}(\mathbf{r}, \mathbf{p})/k_B T) \quad (14)$$

the corresponding thermodynamic potential is

$$-PV/k_B = -\ln Q_{\mu VT}$$

# Thermostats

- NVE is obtained by solving NE.
- NVT can be achieved with the following thermostats:  
Berendsen, Velocity-rescaling, Nose-Hoover.

$$H = \sum_{i=1}^N \frac{\mathbf{p}_i}{2m_i} + U(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) + \frac{p_\xi^2}{2Q} + N_f k T \xi \quad (15)$$

A better approach is Nose-Hoover chain.

- Using general and local thermostats.
- NPT can be simulated with Berendsen and Parrinello-Rahman methods.

# Techniques to speed up simulations

## Serial computation



Figure: Single lane.

# Techniques to speed up simulations

## Parallel computation



Figure: Multiple lanes.

# Techniques to speed up simulations

## Parallel computation



Figure: Multiple lanes can also help to handle larger size problems.

# Techniques to speed up simulations

Parallel computing. Kebnekaise cluster at HPC2N.



**Figure:** Kebnekaise cluster at HPC2N.

# Techniques to speed up simulations

- MPI parallelization
- MPI+OpenMP parallelization
- Domain decomposition scheme
- Multiple communicators

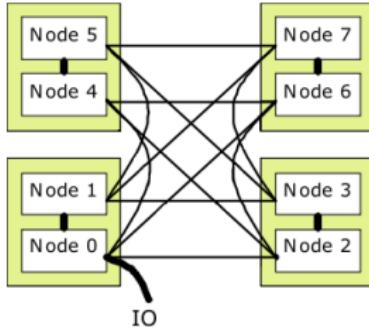


Figure: Nodes (MPI).

```
do i=1,num_particles  
x(i) = x(i) + f(i)*dt  
enddo
```

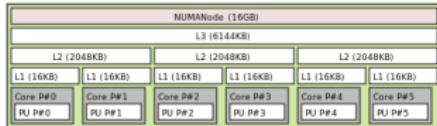


Figure: NUMA machine (OpenMP).

# Coarse-grain simulations

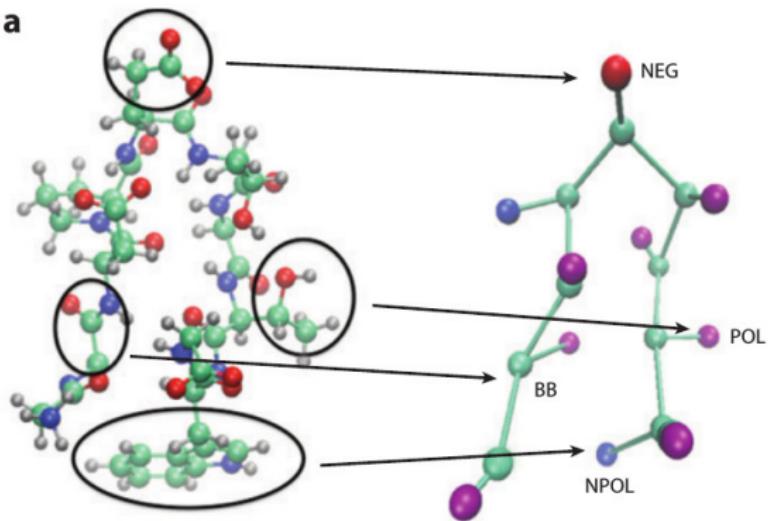
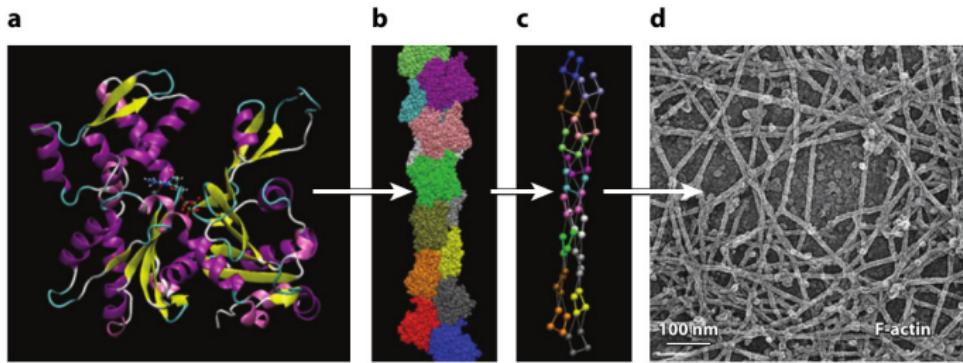


Figure: Reduction of the degrees of freedom [Annu. Rev. Biophys., 42, 73 (2013)].

# Coarse-grain simulations



**Figure 1**

A visual representation of the multiscale challenge for understanding spatiotemporal coupling in biological systems: (a) an atomistic representation of an actin subunit in the monomeric state; (b) an actin filament, made up of many actin subunits; (c) a coarse-grained representation of an actin filament, which reduces the computational cost for simulation; and (d) a mesoscopic cytoskeleton network made up of many individual filaments. (Image in panel d reproduced from <https://science.nichd.nih.gov/confluence/display/sob/Actin+Filament+Networks>.)

**Figure:** Reduction of the degrees of freedom [Annu. Rev. Biophys., 42, 73 (2013)].

# Simulations time scale

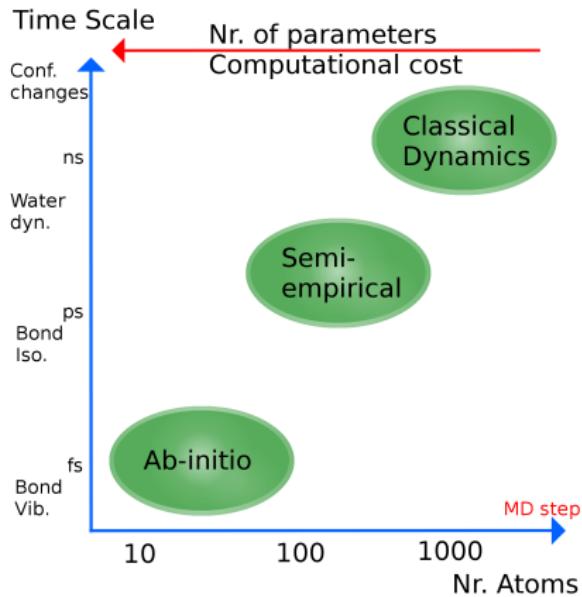


Figure: Accuracy w.r.t. time scale for different modeling approaches.

- NAMD
- GROMACS
- AMBER
- OpenMM
- CHARMM

# Steps for MD

- Setting up the system
- minimization
- solvation
- neutralization
- equilibration
- production
- analysis

# References

- Molecular Modelling, Andrew R. Leach, 2nd. edition
- Computer Simulation of Liquids, M. P. Allen
- Understanding Molecular Simulation, Daan Frenkel and Berend Smit, 2nd. edition