Computational Chemistry - MD Simulations

P. Ojeda-May
pedro.ojeda-may@umu.se

Department of Chemistry/HPC2N,
Umeå University,

901 87, Sweden.

May 2, 2017

Table of contents

Simulations time scale

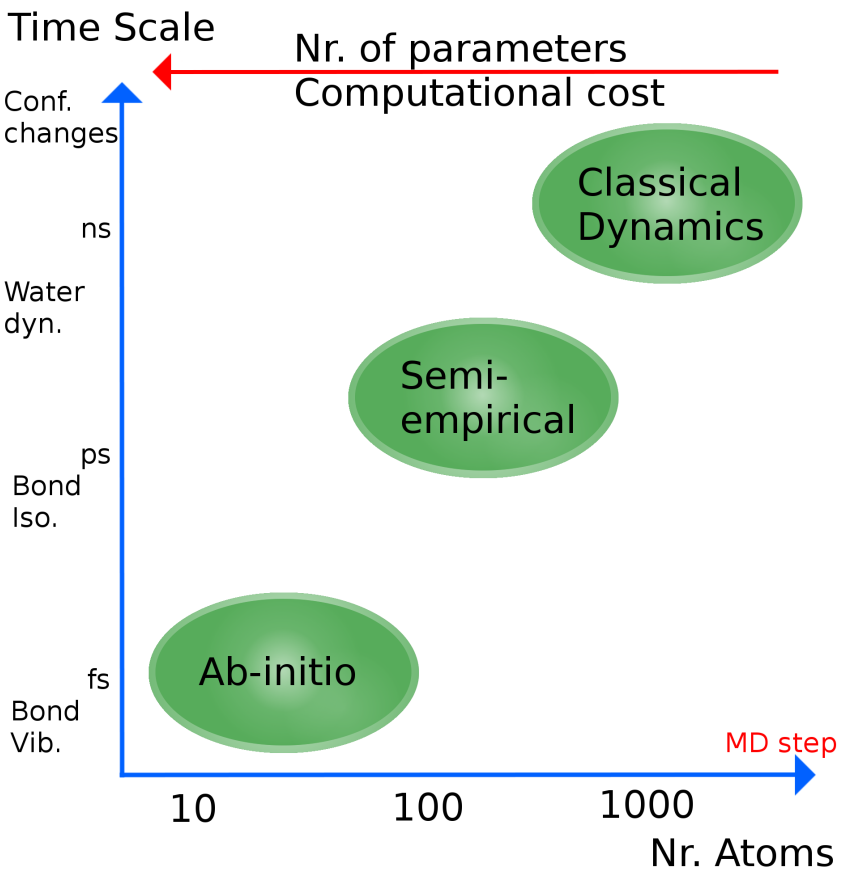


Figure : Accuracy w.r.t. time scale for different modeling approaches.

Early MD simulations

Phase Transition for a Hard Sphere System

B. J. Alder and T. E. Wainwright
University of California Radiation Laboratory, Livermore, California
(Received August 12, 1957)

PHYSICAL REVIEW VOLUME 136, NUMBER 2A 19 O

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, Rehovoth, 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 α helices[1]. 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).

Current MD simulations

The Nobel Prize in Chemistry 2013

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

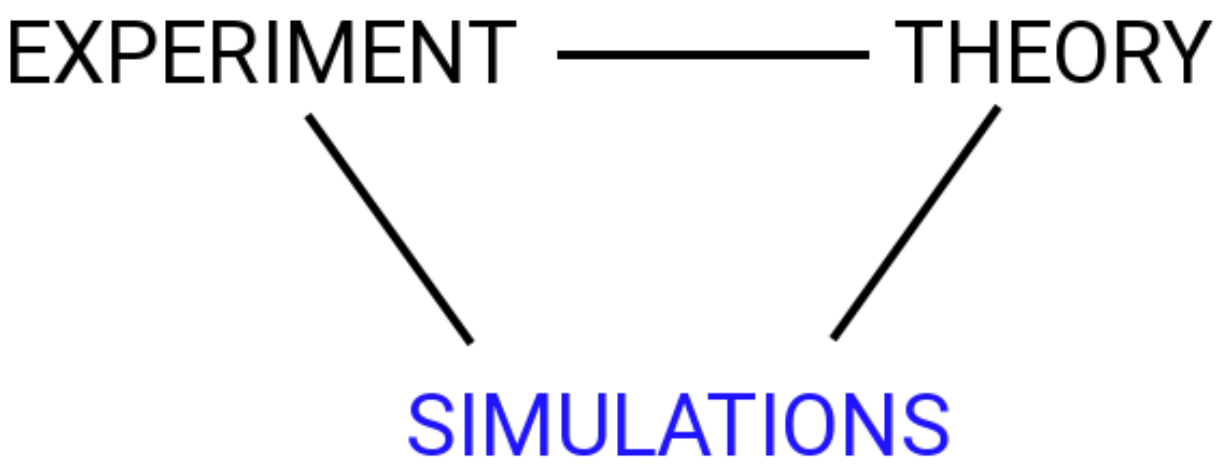
Photo: A. Mahmoud
Michael Levitt
Prize share: 1/3

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 : Source: http://www.nobelprize.org.

Current MD simulations



Application of MD

Proteins

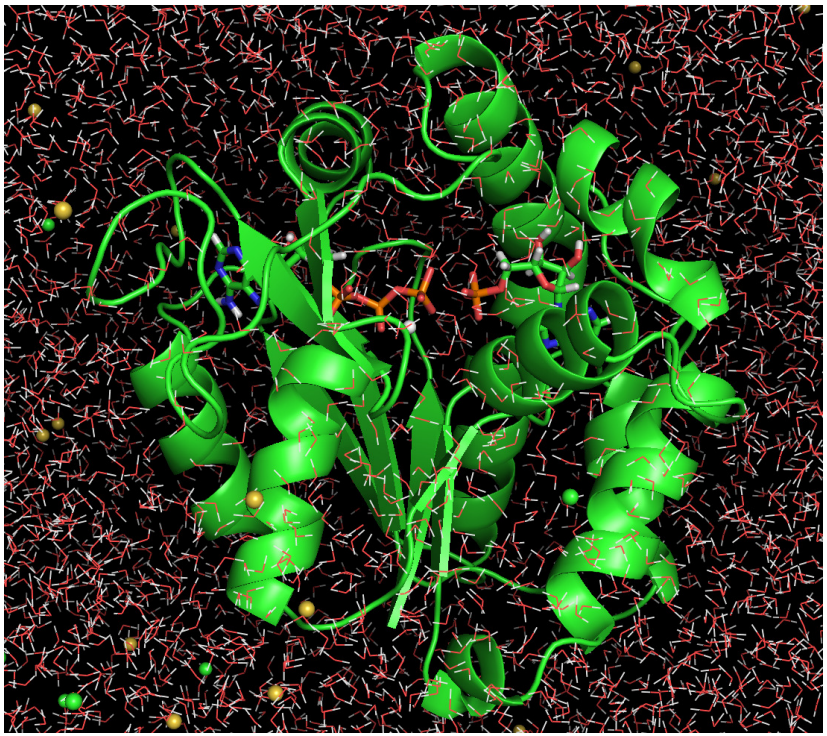


Figure : AdK enzyme in water.

Clays

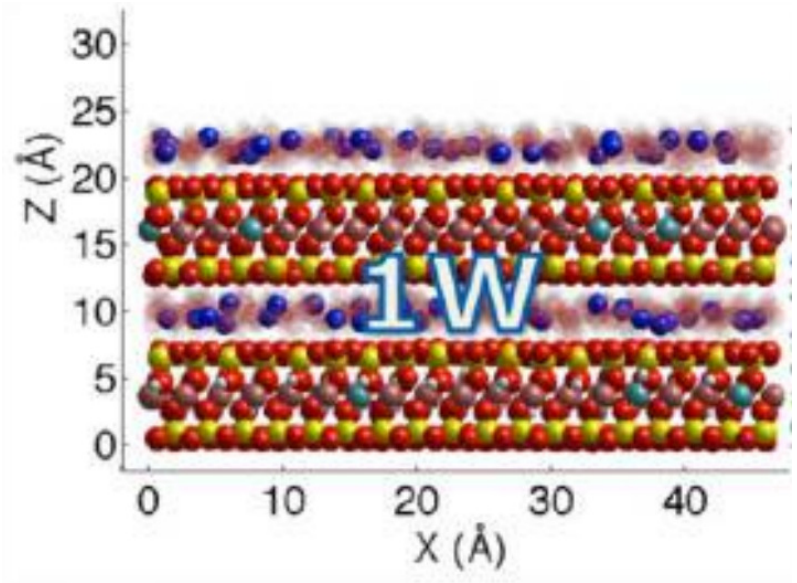


Figure : Clay [JPC C, **118**, 1001 (2014)].

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

ELSEVIER
Construction and Building Materials
journal homepage: www.elsevier.com/locate/conbuildmat

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.

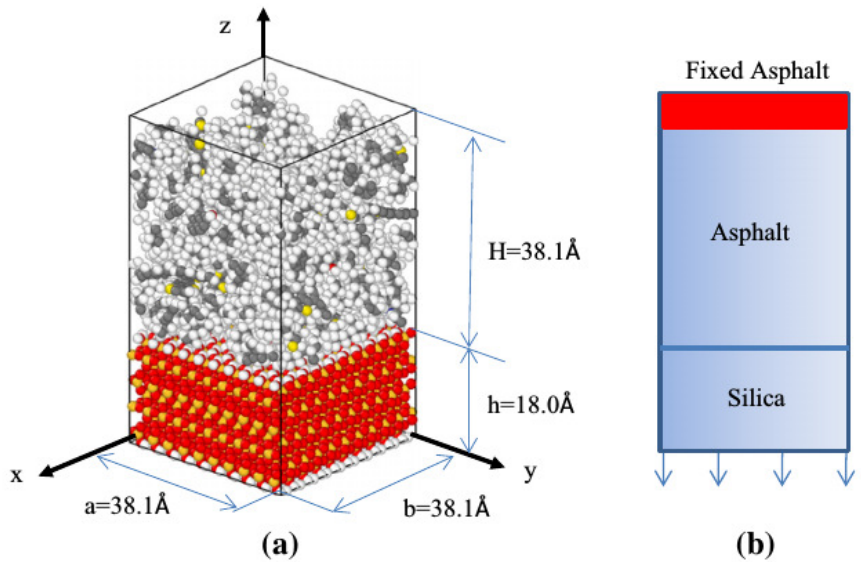


Fig. 3. Representative model for the MD simulations and schematic of tensile simulation.

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

Newton's equation

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

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

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

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

Force fields

Figure : Source: http://www.lpwchem.org/force-field-development/

Force fields

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

Force fields

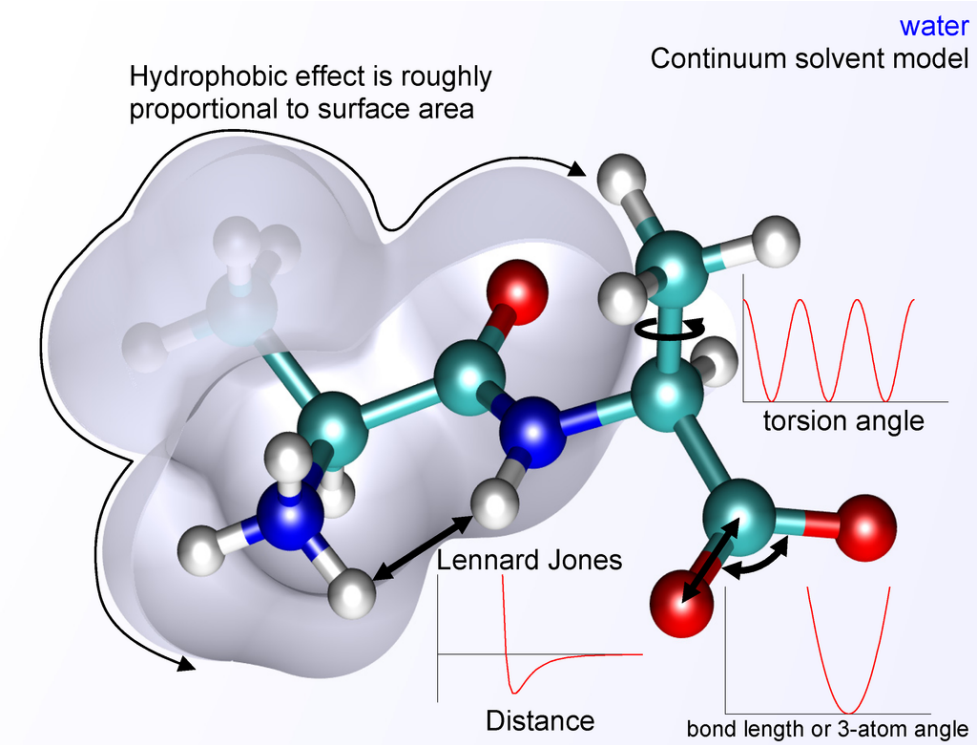


Figure : Energy terms. Source:
https://en.wikipedia.org/wiki/Force_field_(chemistry)

Force fields

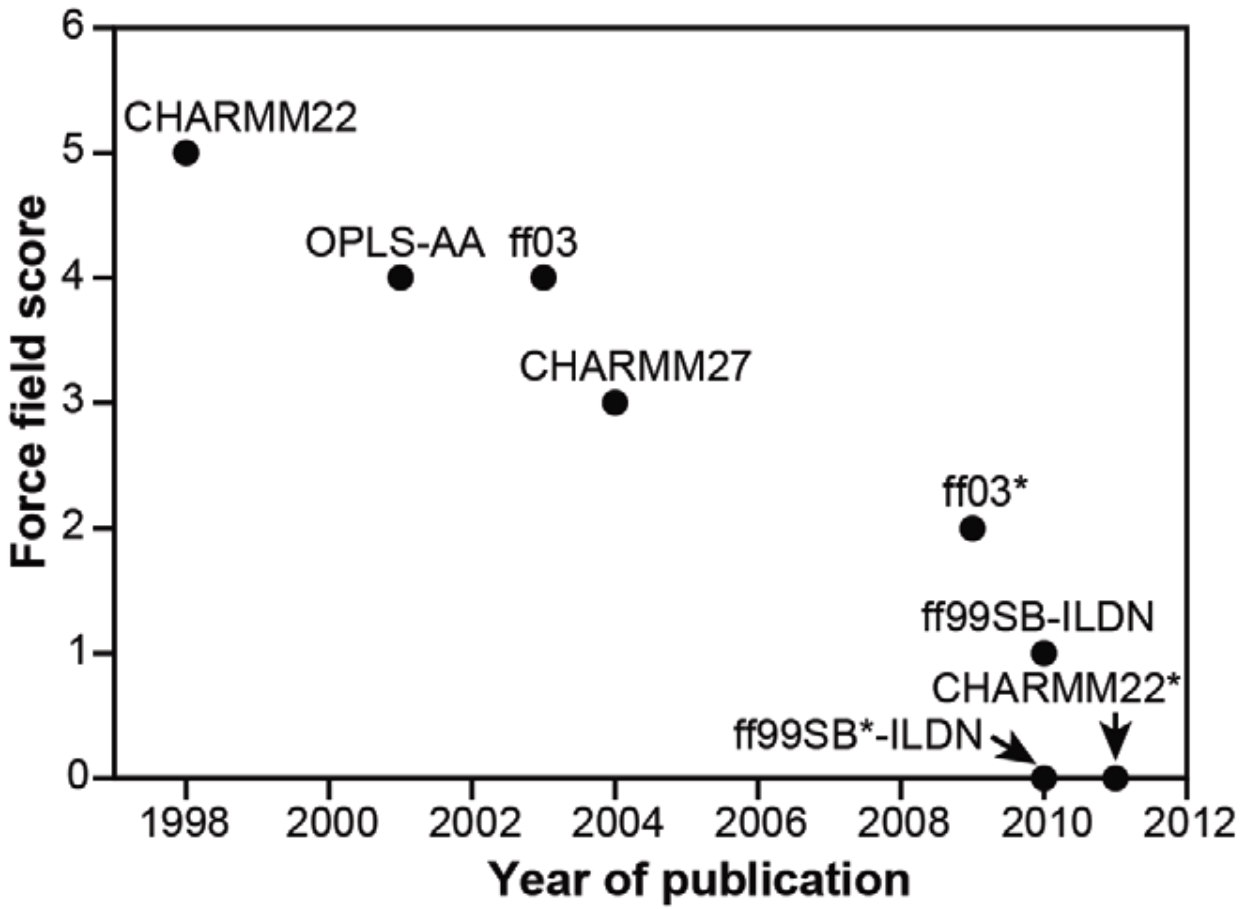


Figure : FF for proteins comparison, PLoS ONE, 7, e32131, (2012).

Force fields: Energy surface

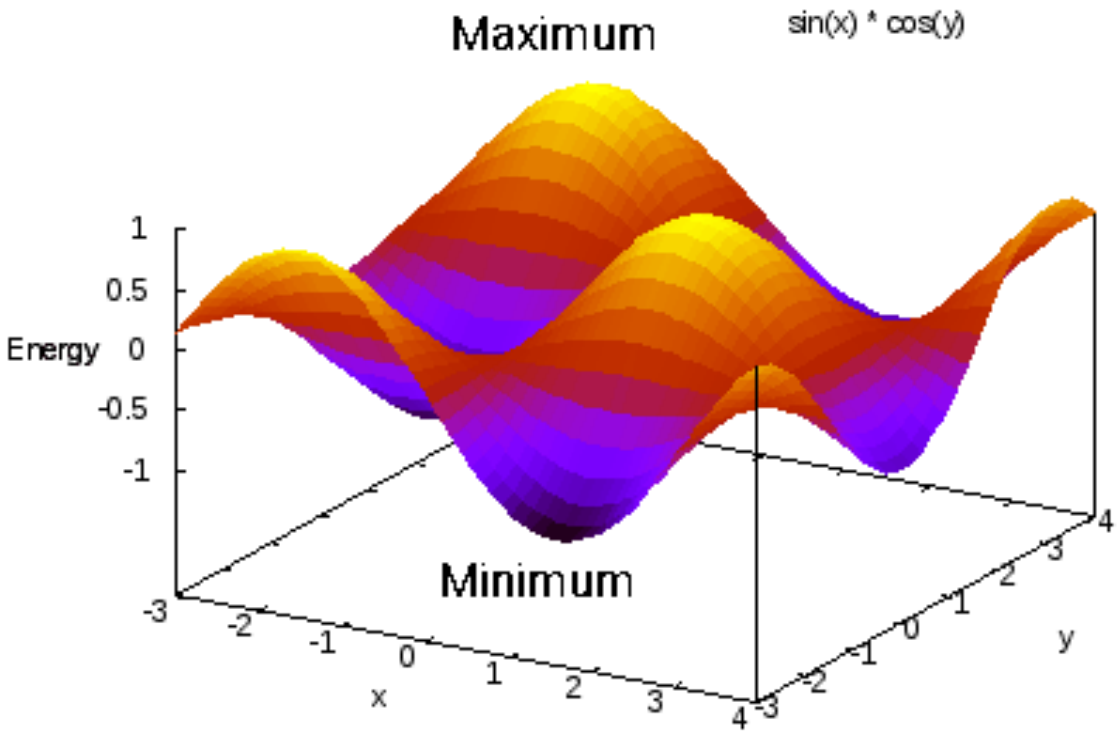


Figure : Energy surface described by $U = sin(x) * cos(x)$

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

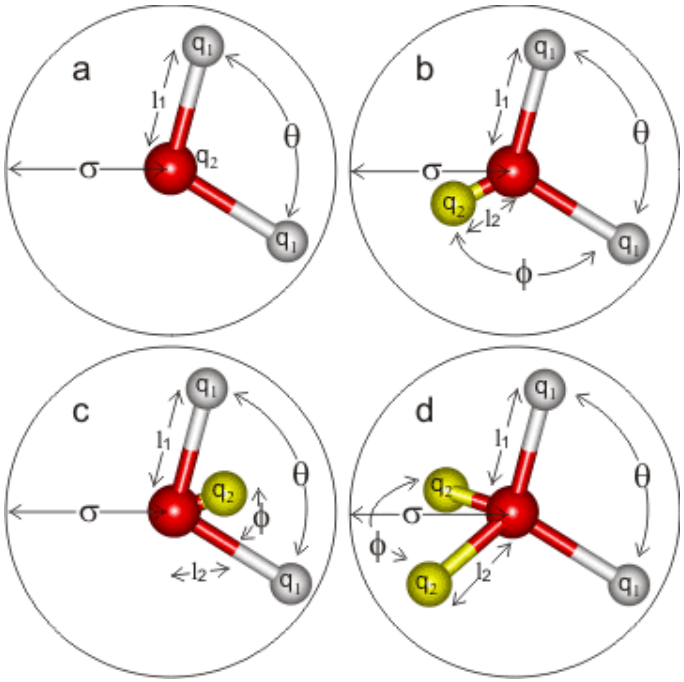


Figure : 3-5 sites water models. Source:
http://www1.lsbu.ac.uk/water/water_models.html

Water models

TABLE 5: Self-diffusion Coefficients ($\times 10^{-9}$ m^2 s^{-1}) for All Water Models Using Two Different Lengths of the Slope of MSD(t) vs t

water model	slope (1.9−9.6 ps)	slope (4.0−20.0 ps)	temperature[c] (K)	D (25 °C)[f]
TIP3P original[a]	5.88[d] (0.10)[e]	5.87[d] (0.09)[e]	301.4[d](1.7)[e]	5.67
TIP3P original[b]	5.59 (0.06)	5.59 (0.08)	297.0 (0.9)	5.65
TIP3P modified[a]	5.92 (0.09)	5.92 (0.11)	301.2 (1.8)	5.73
TIP3P modified[b]	5.83 (0.07)	5.85 (0.08)	299.2 (1.0)	5.78
SPC original[a]	4.39 (0.05)	4.40 (0.06)	301.0 (1.7)	4.22
SPC original[b]	4.22 (0.06)	4.24 (0.08)	298.6 (1.1)	4.20
SPC refined[a]	4.49 (0.08)	4.48 (0.08)	301.0 (1.8)	4.30
SPC refined[b]	4.26 (0.07)	4.24 (0.10)	297.7 (1.2)	4.26
SPC/E original[a]	2.90 (0.06)	2.89 (0.08)	300.4 (1.9)	2.75
SPC/E original[b]	2.78 (0.04)	2.77 (0.06)	298.2 (1.4)	2.76
exptl[8,9]				2.30

[a] Nonbonded list 1 (see Methods). [b] Nonbonded list 2 (see Methods). [c] Temperature of the MD simulation. [d] Mean values. [e] Standard deviations. [f] Self-diffusion coefficients adjusted to 25 °C, using the slope 4.0−20.0 ps.

Figure : See for details: JPC A, **105**, 9954 (2001).

Protein systems

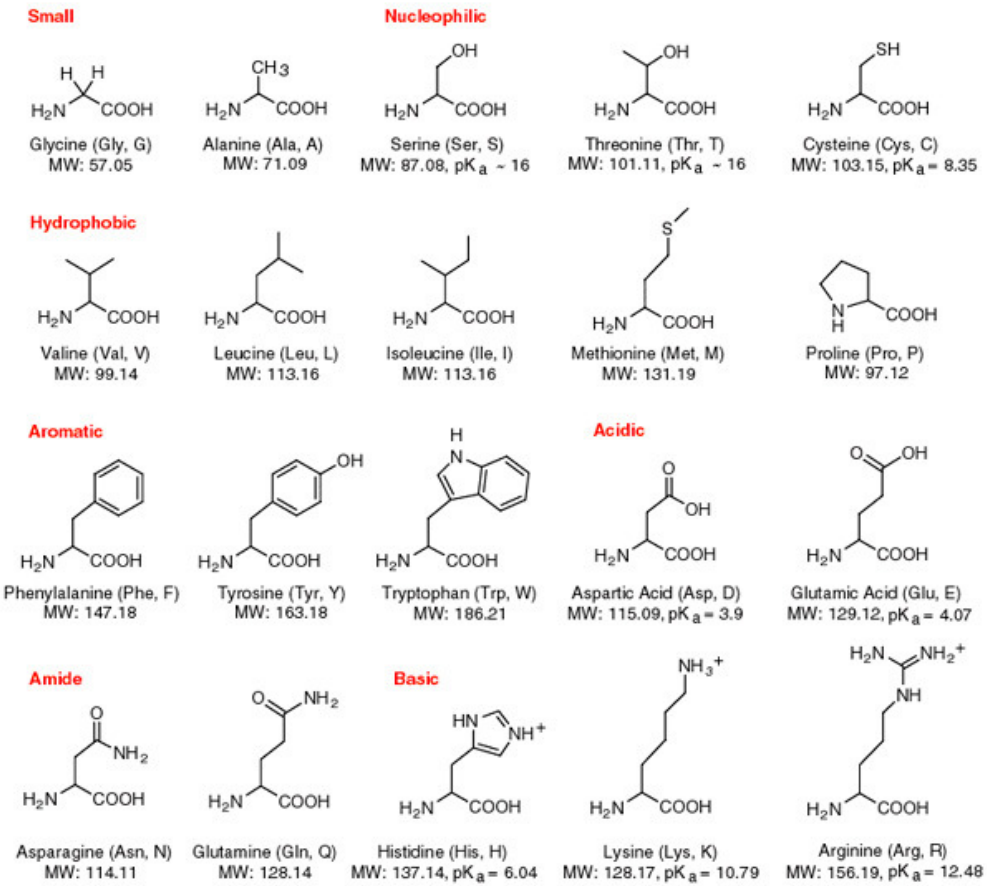


Figure : 20 natural amino acids. Source: goo.gl/YrYvwv

Protein systems

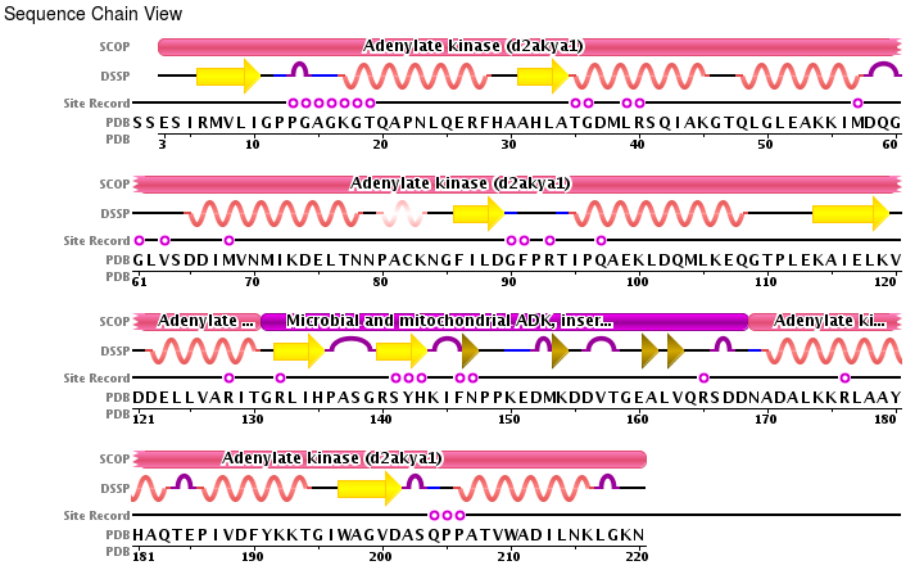


Figure : PDB information of AdK.

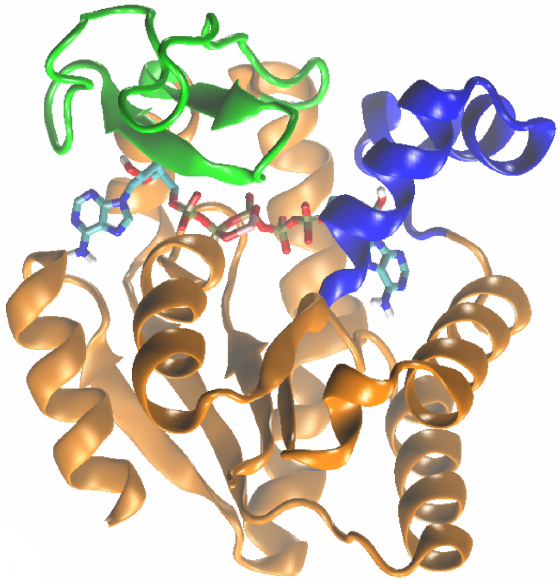


Figure : Structure of yeast AdK.

Periodic boundary conditions (PBC)

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

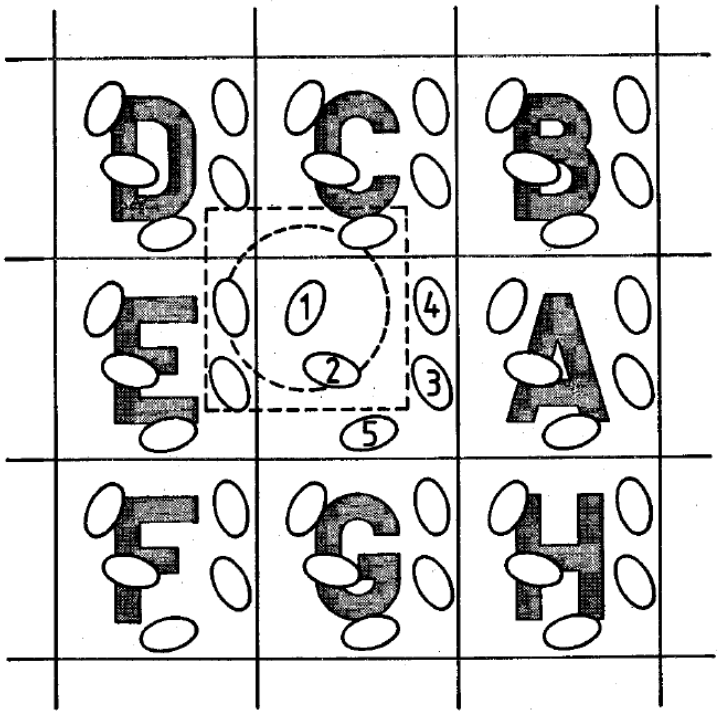


Figure : PBCs and minimum image convention [Allen & Tildesley, Comp. Sim. of Liquids]

Electrostatic interactions: Ewald method

The elecrostatic energy for a periodic system can be written as[1],

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

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.

[1] Adv. Polym. Sci., **185** , 59 (2005)

Electrostatic interactions: Ewald method

The elecrostatic energy for a periodic system can be written as[1],

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

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.
The potential is splitted such that,

$$\frac{1}{r} = \frac{f(r)}{r} + \frac{1 - f(r)}{r} \tag{6}$$

[1] Adv. Polym. Sci., **185** , 59 (2005)

Electrostatic interactions: Ewald method

The elecrostatic energy for a periodic system can be written as[1],

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

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.
givig rise to the total energy:

$$E = E^{(r)} + E^{(k)} + E^{(s)} + E^{(d)} \tag{6}$$

[1] Adv. Polym. Sci., **185** , 59 (2005)

Electrostatic interactions

$$E^{(r)} = \frac{1}{2} \sum_{m \in \mathbb{Z}^3}^{\infty} \sum_{i,j=1}^{N}{}' q_i q_j \frac{\mathrm{erfc}(\alpha |\mathbf{r}_{ij} + \mathbf{m}L|)}{|\mathbf{r}_{ij} + \mathbf{m}L|} \tag{7}$$

$$E^{(k)} = \frac{1}{2V} \sum_{k \neq 0} \frac{4\pi}{k^2} e^{k^2/4\alpha^2} |\tilde{\rho}(\mathbf{k})|^2 \tag{8}$$

$$E^{(s)} = -\frac{\alpha}{\sqrt{\pi}} \sum_i q_i^2 \tag{9}$$

$$E^{(d)} = \frac{2\pi}{(1+2\epsilon')V} (\sum_i q_i \mathbf{r}_i)^2 \tag{10}$$

Integration of Newton's equation

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

$$\mathbf{F} = m\mathbf{a} = -\nabla U \qquad \text{Newton's Law} \tag{11}$$

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

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \delta t \mathbf{v}(t+\frac{1}{2}\delta t) \tag{12}$$

$$\mathbf{v}(t+\frac{1}{2}\delta t) = \mathbf{v}(t-\frac{1}{2}\delta t) + \delta t \mathbf{a} \tag{13}$$

velocities are updated according to,

$$\mathbf{v}(t) = \frac{1}{2}\left(\mathbf{v}(t+\frac{1}{2}\delta t) + \mathbf{v}(t-\frac{1}{2}\delta t)\right) \tag{14}$$

Constraints

Collision of two diatomic molecules

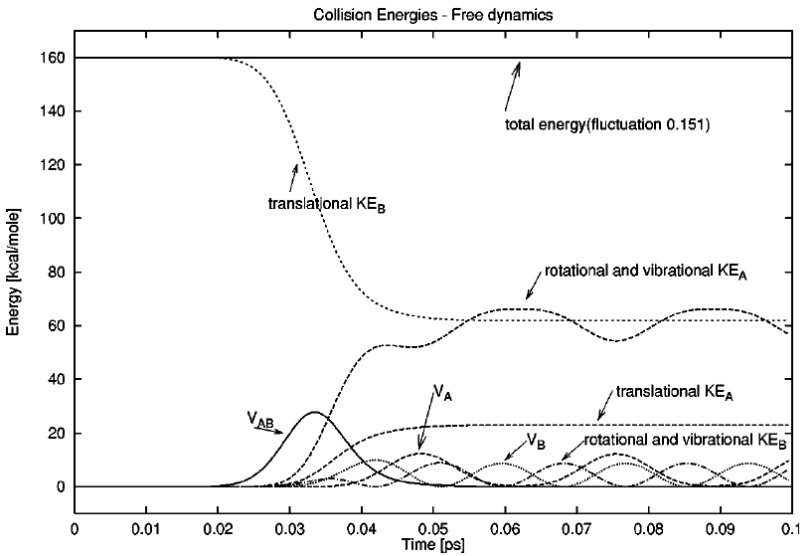


Figure : Free collision.

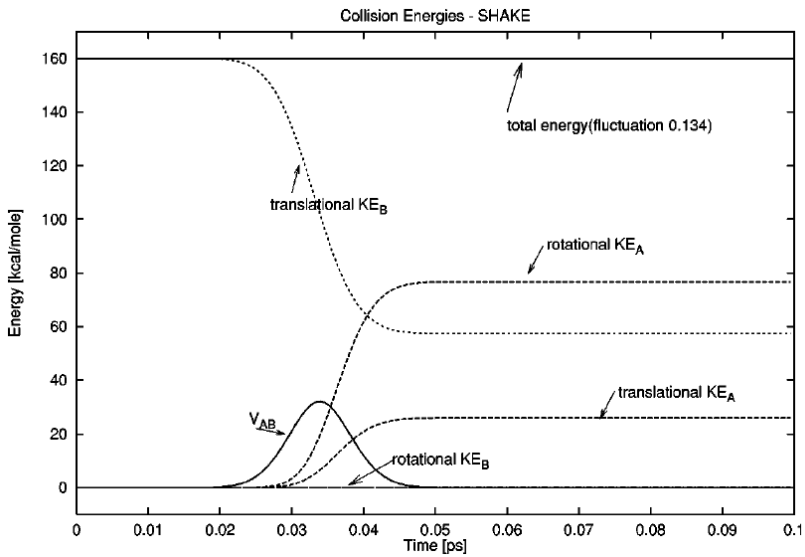


Figure : SHAKE constraint.

See JCP, **112**, 7919 (2000)

Constraints

Modern approaches to deal with constraints

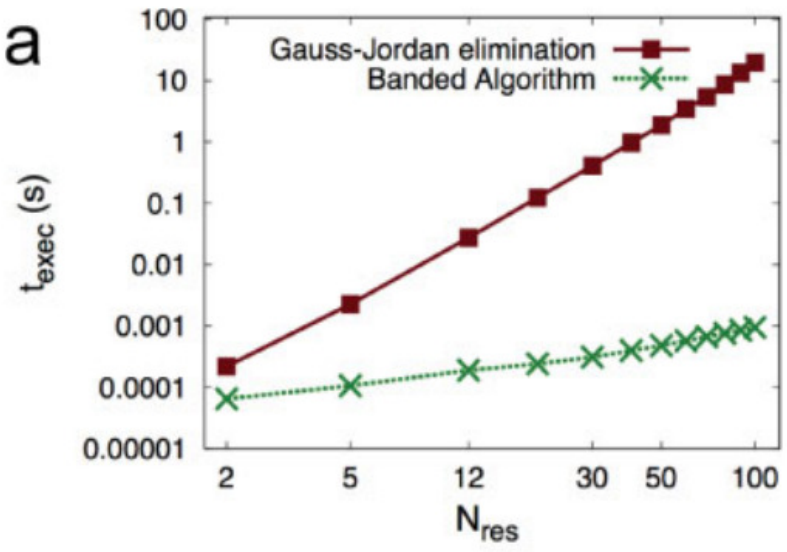


Figure : ILVES method.

See JCC, **32**, 3039 (2011)

Techniques to speedup simulations

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

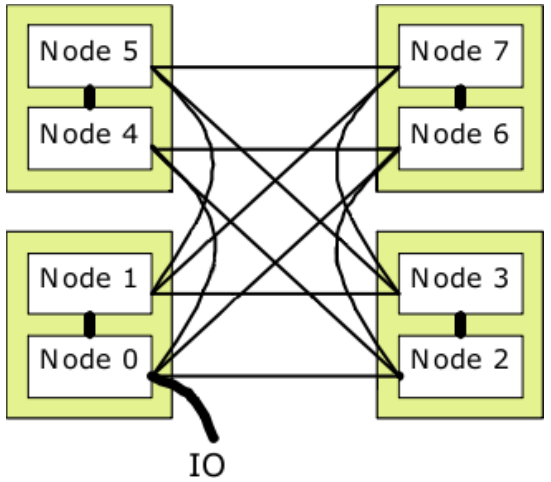


Figure : Nodes (MPI).

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

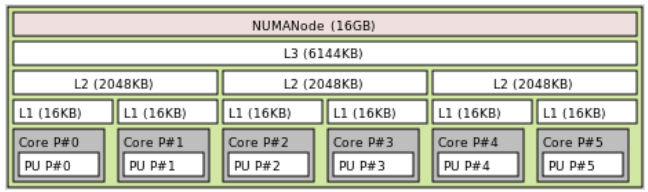


Figure : NUMA machine (OpenMP).

Ergodicity

$$\begin{aligned} \mathcal{A}_{obs} &= < \mathcal{A} >_{\text{time}} \\ &= < \mathcal{A}(\Gamma(t)) >_{\text{time}} \\ &= \lim_{t_{\text{obs}} \to \infty} \int_0^{t_{\text{obs}}} \mathcal{A}(\Gamma(t)) dt \end{aligned} \tag{15}$$

Figure : Coffee cup.

Statistical ensembles

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

$$Q_{NVE} = \frac{1}{N!}\frac{1}{h^{3N}}\int \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{p}\delta(\mathcal{H}(\mathbf{r},\mathbf{p}) - E) \qquad (16)$$

  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,

$$Q_{NVT} = \frac{1}{N!}\frac{1}{h^{3N}}\int \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{p}\exp(-\mathcal{H}(\mathbf{r},\mathbf{p})/k_B T) \qquad (17)$$

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

Statistical ensembles

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

$$Q_{NPT} = \frac{1}{N!}\frac{1}{h^{3N}}\frac{1}{V_0}\int \mathrm{d}V \int \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{p}\exp(-(\mathcal{H}(\mathbf{r},\mathbf{p})+PV)/k_B T) \tag{18}$$

  the corresponding thermodynamic potential is
  $G/k_B = -\ln Q_{NPT}$

- Grand-canonical ensemble (μVT) partition function is,

$$Q_{\mu VT} = \sum_N \frac{1}{N!}\frac{1}{h^{3N}}\exp(\mu N/k_B T)\int \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{p}\exp(-\mathcal{H}(\mathbf{r},\mathbf{p})/k_B T) \tag{19}$$

  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.

$$H = \sum_{i=1}^{N} \frac{\mathbf{p}_i}{2m_i} + U(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) + \frac{p_\xi^2}{2Q} + N_f kT\xi \qquad (20)$$

  A better approach is Nose-Hoover chain.
- Using general and local thermostats.
- NPT can be simulated with Berendsen and Parrinello-Rahman methods.

Accelerated MD simulations

The original potential energy surface $V(\mathbf{r})$ is modified according to,

$$V^*(\mathbf{r}) = \begin{cases} V(\mathbf{r}), & V(\mathbf{r}) \geq E, \\ V(\mathbf{r}) + \Delta V(\mathbf{r}) & V(\mathbf{r}) < E. \end{cases} \tag{21}$$

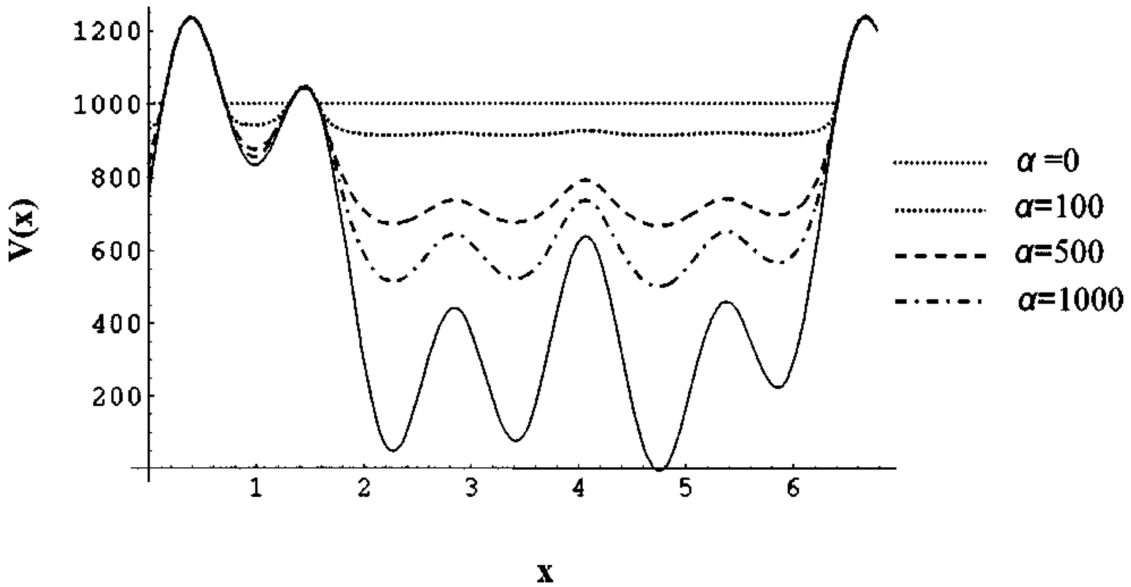


Figure : Modified potential energy surface [JCP, **120**, 11919 (2004)].

Accelerated MD simulations

the biasing term is,

$$\Delta V(\mathbf{r}) = \frac{(E - V(\mathbf{r}))^2}{\alpha + (E - V(\mathbf{r}))} \tag{22}$$

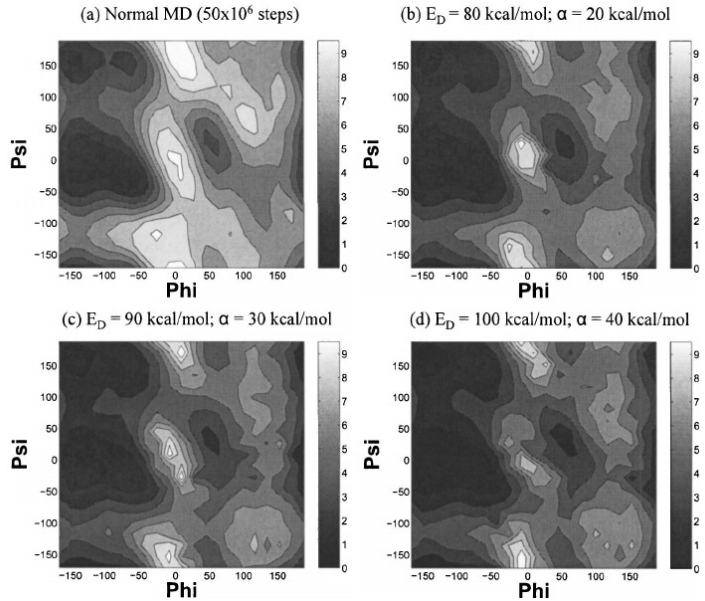


Figure : Free energy landscape of Alanine dipeptide [JCP, **120**, 11919 (2004)].

Umbrella sampling (US) simulations

The potential energy is modified as follows JCP, **23**, 187 (1977):

$$E^b(r) = E^u(r) + w_i(\xi)$$

with $w_i(\xi) = K/2(\xi - \xi_i^{ref})^2$

For each window the free energy is given by,

$$A_i(\xi) = -(1/\beta) \ln P_i^b(\xi) - w_i(\xi) + F_i$$

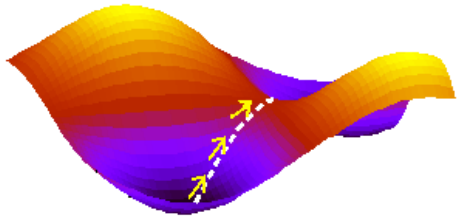


Figure : Potential energy surface.

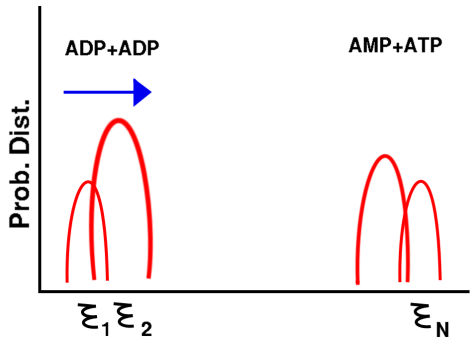


Figure : Probability histograms.

String method (SM) simulations

Define a set of collective variables z_j and effective forces as follows

$$\frac{k}{T}\int_0^T (z_j - \theta_j(t))dt \sim \frac{\partial F(z)}{\partial z_j}$$

The free energy along the string is computed by PRB, **66**, 052301 (2002),

$$F(z(\alpha)) - F(z(0)) = \int_0^\alpha \sum_{i=1}^{N} \frac{dz_i(\alpha')}{d\alpha'} \frac{\partial F(z(\alpha'))}{\partial z_i} d\alpha'$$

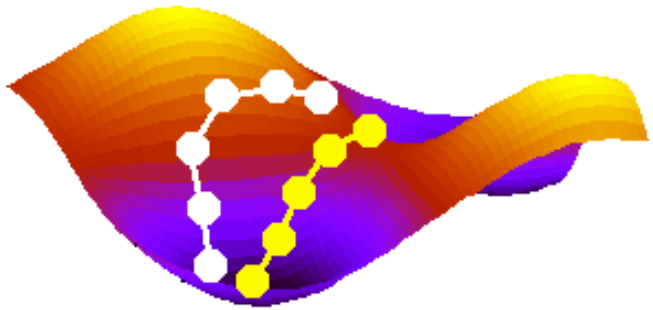


Figure : Free energy surface.

String method (SM) simulations

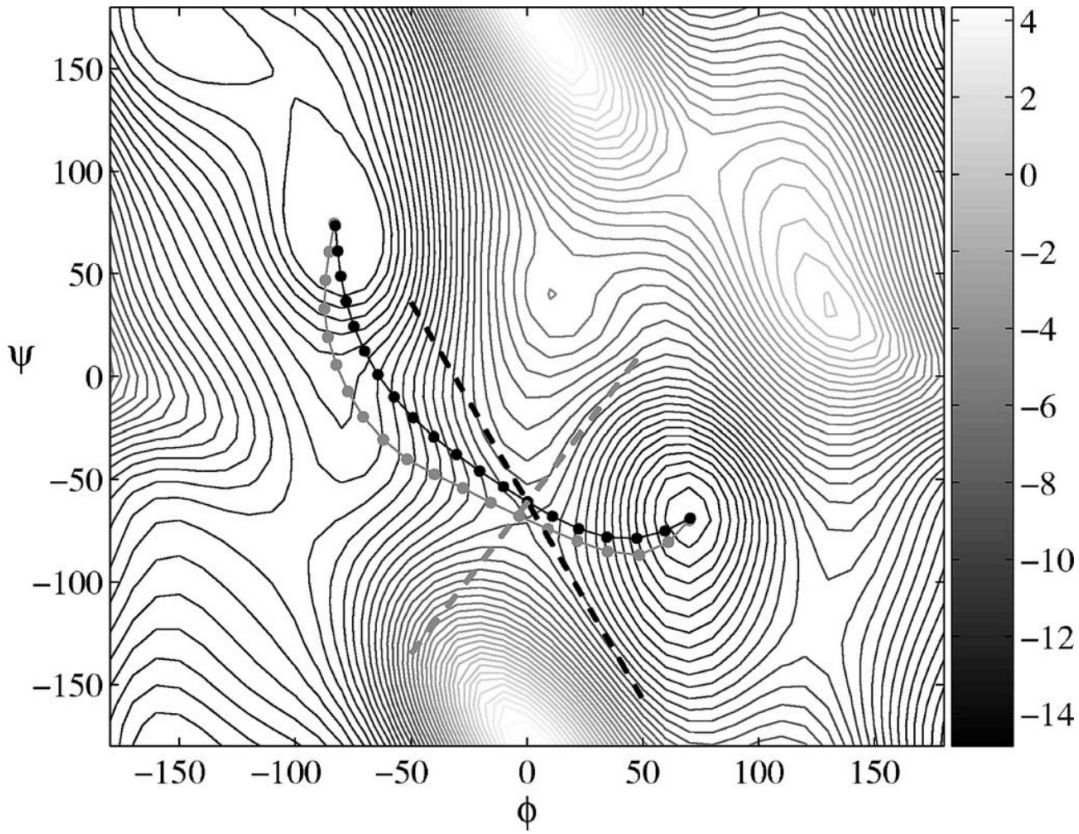


Figure : Free energy surface of Alanine dipeptide.

Coarse-grain simulations

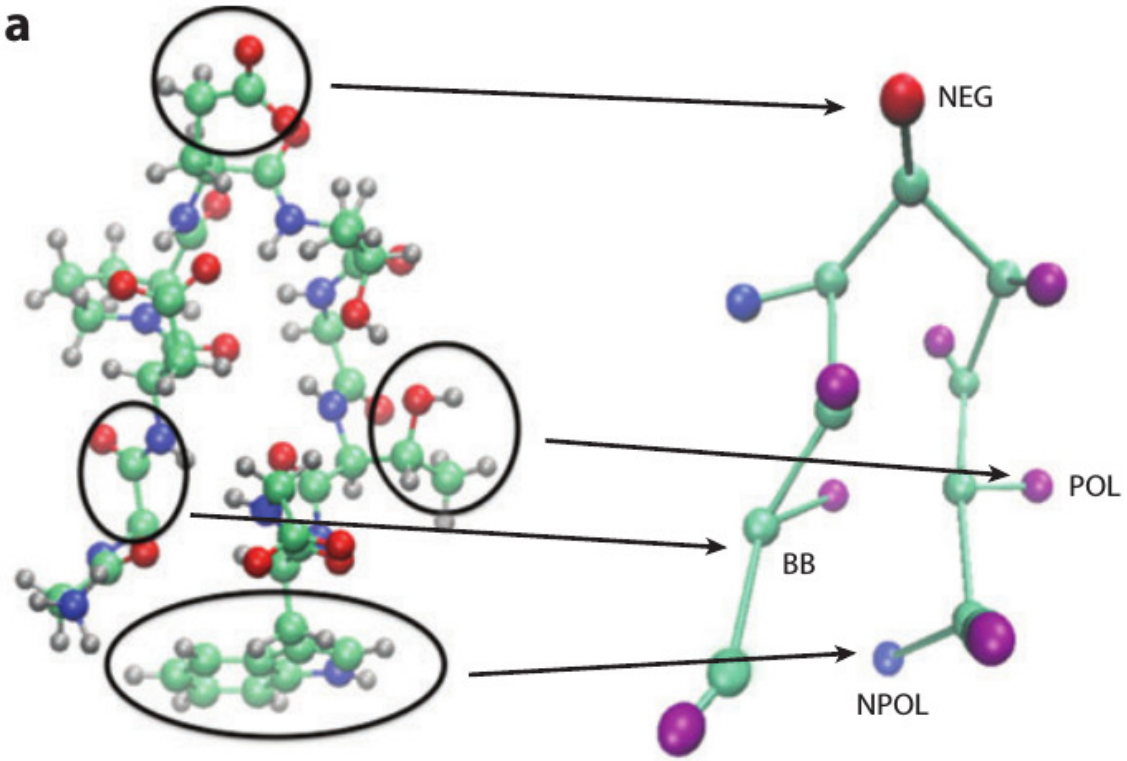


Figure : Reduction of the degrees of freedom [Annu. Rev. Biophys., **42**, 73 (2013)].

Coarse-grain simulations

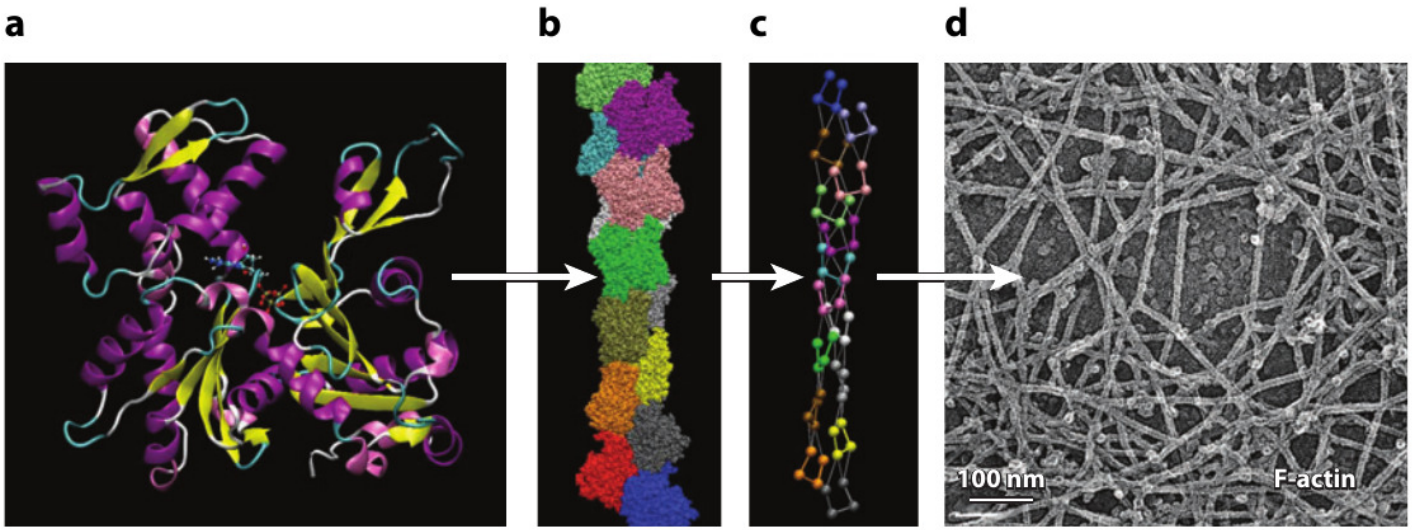


Figure 1

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)].

Alchemical simulations

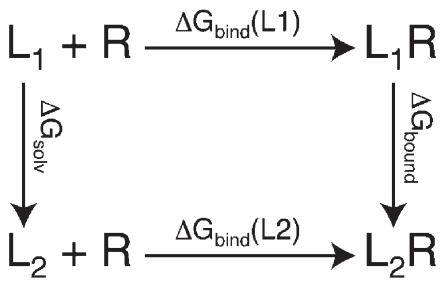


Figure : Thermodynamic cycle for binding of two protein ligands L_1 and L_2, [JCC, **30**, 1692 (2009)].

$$\Delta\Delta G^{bind}_{L_i \to L_j} = \Delta G^{bind}_{L_j} - \Delta G^{bind}_{L_i} = \Delta G^{prot}_{RL_i \to RL_j} - \Delta G^{solv}_{L_i \to L_j} \quad (23)$$

The Hamiltonian is modified according to,

$$H = T_x + (1-\lambda)V_0 + \lambda V_1 \quad (24)$$

Alchemical simulations

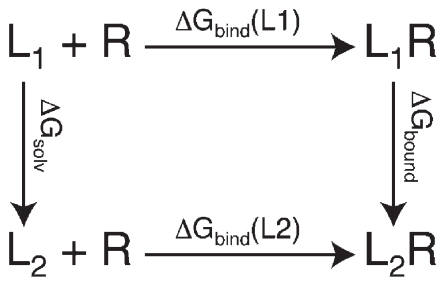


Figure : Thermodynamic cycle for binding of two protein ligands L_1 and L_2, [JCC, **30**, 1692 (2009)].

The free energy difference going from $\lambda = 0$ to $\lambda = 1$ is,

$$\Delta G_{\lambda=0\rightarrow\lambda=1} = \sum_{\lambda=0}^{1} -\frac{1}{\beta} \ln \left\langle \exp\left(-\beta(H_{(\lambda+\delta\lambda)} - H_{(\lambda)})\right)\right\rangle \qquad (25)$$

Installing CHARMM on Kebnekaise

```
Load the modules
module load GCC/5.4.0-2.26  OpenMPI/1.10.3

go to the charmm folder and type:
./install.com gnu M

at the end of the installation one gets the message:
 install.com> CHARMM Installation is completed.
                The CHARMM executable is /pfs/nobackup/home/
```

Running CHARMM on Kebnekaise

```
#!/bin/bash
#SBATCH -A project-ID
#SBATCH -N 1
#SBATCH --time=01:00:00
#SBATCH --output=job_str.out
#SBATCH --error=job_str.err
#SBATCH --exclusive
#SBATCH --mail-type=END

module load GCC/5.4.0-2.26  OpenMPI/1.10.3

mpirun -np 28 /home/u/user/pfs/CHARMM_tutorial/charmm/exec/
 -i step4_equilibration.inp > out_equilibration.dat
mpirun -np 28 /home/u/user/pfs/CHARMM_tutorial/charmm/exec/
```

CHARMM

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

CHARMM files

- *.pdb (coordinates)
- CHARMM parameter files
- *.inp (input file for simulation)

Performance of CHARMM

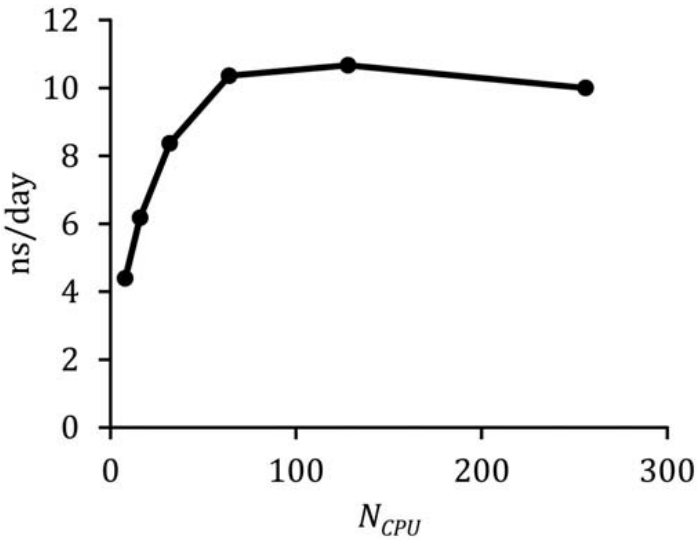


Figure : Old CHARMM version.

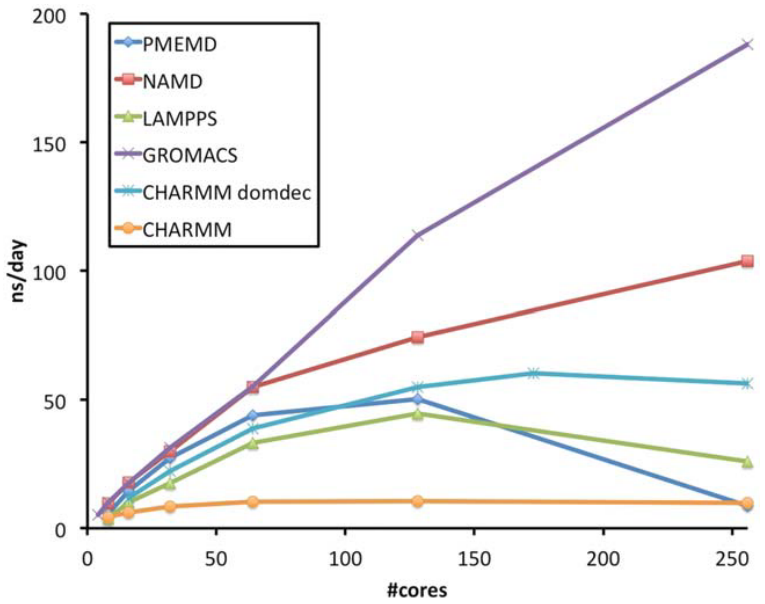


Figure : Comparison of CHARMM with other novel software.

System of 19,609 atoms. JCC, **35**, 406-413 (2014)

CHARMM support

- https://www.charmm.org/ubbthreads/
- contact HPC2N support