

Computational Chemistry - MD Simulations

P. Ojeda-May

Department of Chemistry/HPC2N,
Umeå University,



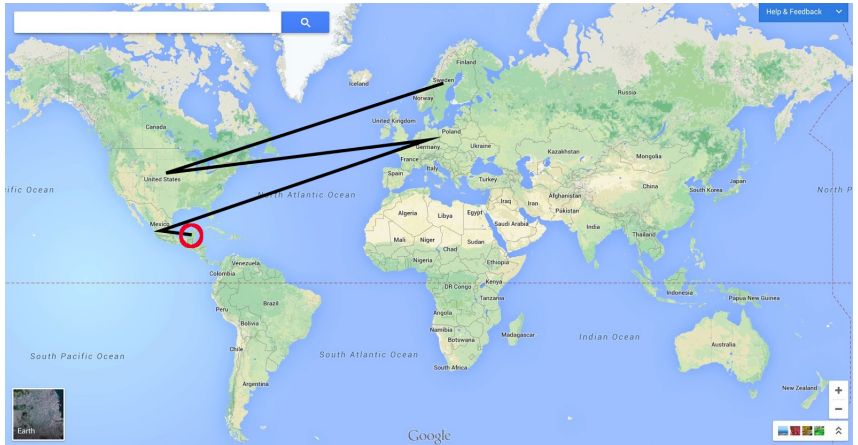
901 87, Sweden.

June 17, 2016

Table of contents

- 1 Basics on MD simulations
 - Force field
 - Ensembles
 - Beyond classical MD
 - Accelerated MD
 - Umbrella sampling
 - String method
 - Coarse graining
 - Alchemical method
- 2 Abisko and Kebnekeise
- 3 Using GROMACS at HPC2N
 - Setting up the system, minimization, solvation, equilibration, production and analysis.

Background



Background

- Physicist currently working on Chemistry.
- Experience with Supercomputers and HPC Clusters.

Background

- Physicist currently working on Chemistry.
- Experience with Supercomputers and HPC Clusters.
- Code developer of Chemistry at Harvard Molecular Mechanics (CHARMM) package:
 - Carr-Parrinello Molecular Dynamics
 - Multiple time steps algorithms
 - Modified Neglect of Diatomic Overlap (MNDO) QM/MM approach
 - Hybrid MPI/OpenMP approaches for QM/MM simulations
 - QM/MM-Cutoff methods.

Background

- Physicist currently working on Chemistry.
- Experience with Supercomputers and HPC Clusters.
- Code developer of Chemistry at Harvard Molecular Mechanics (CHARMM) package:
 - Carr-Parrinello Molecular Dynamics
 - Multiple time steps algorithms
 - Modified Neglect of Diatomic Overlap (MNDO) QM/MM approach
 - Hybrid MPI/OpenMP approaches for QM/MM simulations
 - QM/MM-Cutoff methods.

Simulations time scale

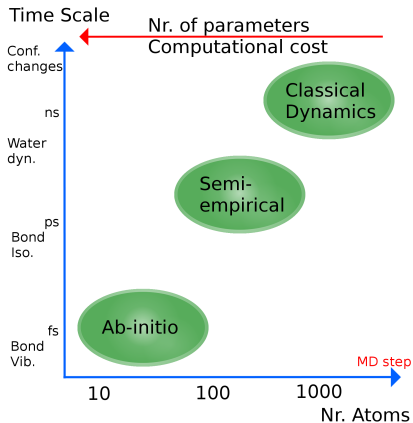


Figure : Accuracy w.r.t. time scale for different modeling approaches.

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 0

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'. 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

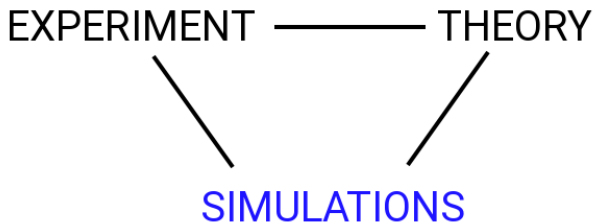


Photo: A. Mahmoud
Arieh Warshel
Prize share: 1/3

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 : Taken from: <http://www.nobelprize.org>.

Current MD simulations



Application of Parallel algorithms

Proteins

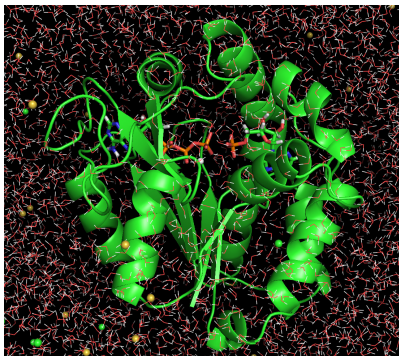


Figure : AdK enzyme in water.

Clays

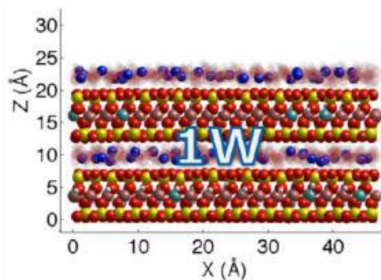


Figure : Clay [JPC C, **118**, 1001 (2014)].

Application of Parallel algorithms

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 Parallel algorithms



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.

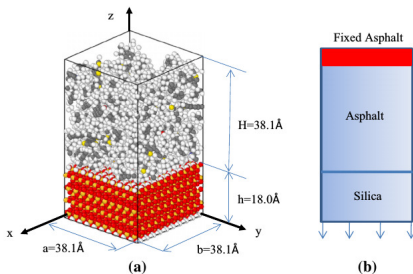


Fig. 3. Representative model for the MD simulations and schematic of tensile simulation.

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

Newton's equation

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

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

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

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

Force fields



Figure : Taken from: <http://www.lpwchem.org/force-field-development/>

Force fields

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

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

Force fields: Energy surface

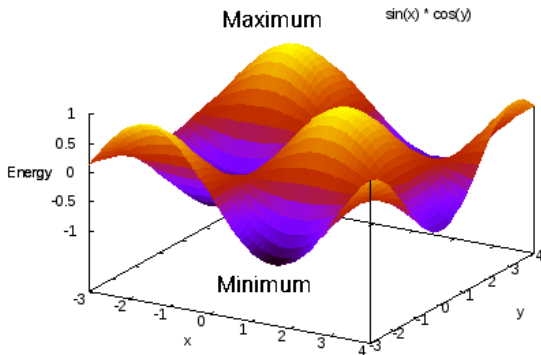


Figure : Energy surface described by $V = \sin(x) * \cos(x)$

Water models

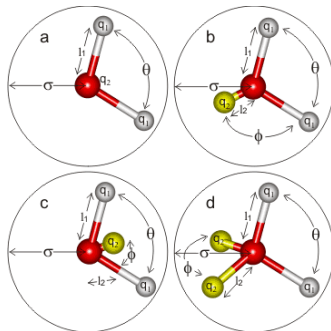


Figure : 3-5 sites water models. Taken from:
http://www1.lsbu.ac.uk/water/water_models.html

Water models

TABLE 5: Self-diffusion Coefficients ($\times 10^{-9} \text{ m}^2 \text{ s}^{-1}$) 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

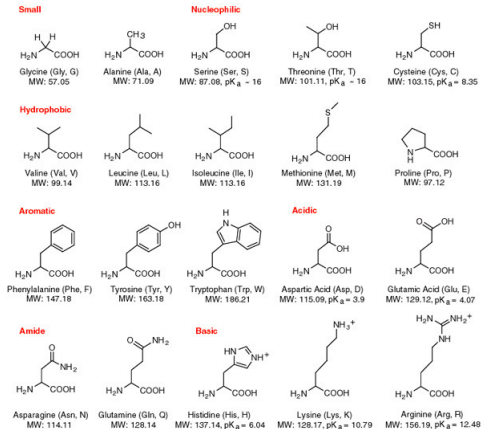


Figure : 20 natural amino acids. Taken from: goo.gl/YrYvww

Protein systems

Sequence Chain View

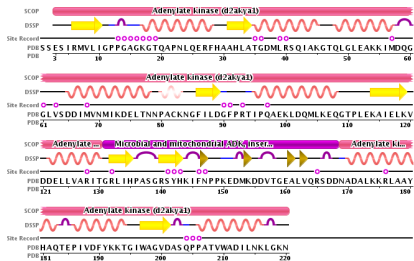


Figure : PDB information of AdK.

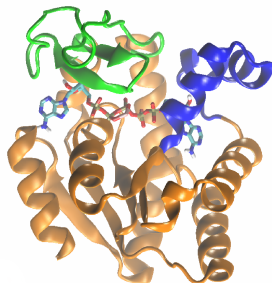


Figure : Structure of yeast AdK.

Periodic boundary conditions (PBC)

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

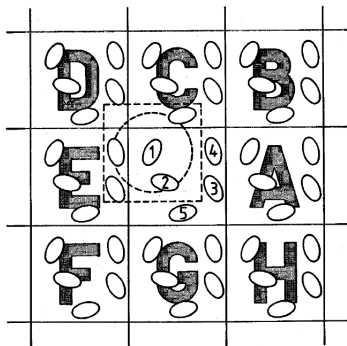


Figure : PBCs and minimum image convention [Allen & Tildesley, Comp. Sim. of Liquids]

Electrostatic interactions: Ewald method

The electrostatic energy for a periodic system can be written as¹,

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

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 .

¹Adv. Polym. Sci., **185**, 59 (2005)

Electrostatic interactions: Ewald method

The electrostatic energy for a periodic system can be written as¹,

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

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 .

The potential is splitted such that,

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

¹Adv. Polym. Sci., **185**, 59 (2005)

Electrostatic interactions: Ewald method

The electrostatic energy for a periodic system can be written as¹,

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

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 .

giving rise to the total energy:

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

¹Adv. Polym. Sci., **185**, 59 (2005)

Electrostatic interactions

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

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

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

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

Electrostatic interactions: Cutoff methods

The electrostatic energy for a periodic system can be written as²,

$$E = \frac{1}{2} \sum_{\mathbf{m} \in \mathbb{Z}^3} \sum_{i,j=1}^N \prime \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{m}L|} \quad (11)$$

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 $\mathbf{m} = 0$.

Can we truncate the interactions up to $r = R_c$?

$$E = \frac{1}{2} \sum_i^N \sum_j^{\mathbf{r}_{ij} < R_c} \frac{1}{r_{ij}} + \Phi \quad (12)$$

²Adv. Polym. Sci., **185**, 59 (2005)

Electrostatic interactions: NaCl lattice

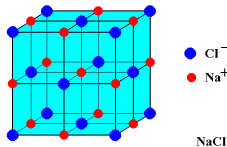


Figure : NaCl lattice.
(source:goo.gl/Fa7tcL)

$$E_i(R_c) = \sum_{\substack{j \neq i \\ (r_{ij} < R_c)}} \frac{q_i q_j}{r_{ij}} \quad (13)$$

$$E^{Mad} = -3.495129...q^2/a$$

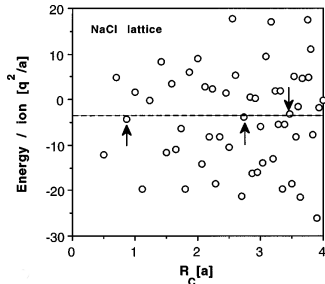


Figure : Single ion energy for NaCl lattice. (Wolf et al., JCP, **110**, 8256 (1999))

Electrostatic interactions: NaCl lattice

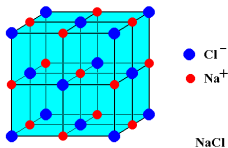


Figure : NaCl lattice.
(source:goo.gl/Fa7tcL)

$$E_i(R_c) = \sum_{\substack{j \neq i \\ (r_{ij} < R_c)}} \frac{q_i q_j}{r_{ij}} \quad (13)$$

$$E^{Mad} = -3.495129...q^2/a$$

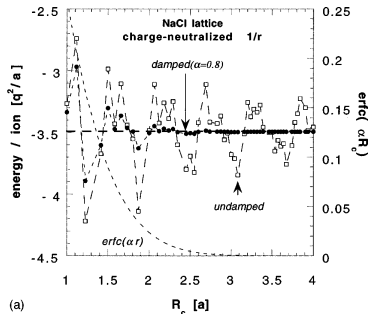


Figure : Energy convergence upon charge neutralization. (Wolf et al., JCP, 110, 8256 (1999))

Isotropic Periodic Sum method: NaCl lattice

$$\epsilon_{ij}^{\text{IPS}}(r_{ij}) = \begin{cases} \epsilon_{ij}(r_{ij}) + \phi_{ij}(r_{ij}) & \text{if } r_{ij} \leq R_c \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

ϵ_{ij} is the Coulombic term and ϕ_{ij} is the long-range IPS correction whose operational expression is given by,

$$\phi_{ij}(r_{ij}) = \frac{q_i q_j}{R_c} \left[\sum_{k=1}^6 b_{2k} \left(\frac{r_{ij}}{R_c} \right)^{2k} \right] \quad (15)$$

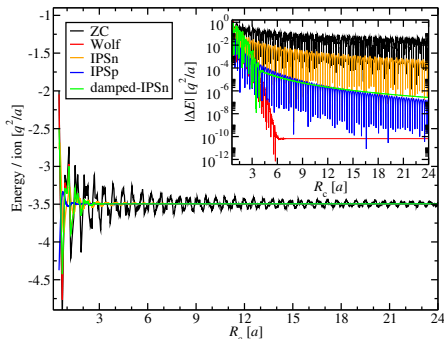


Figure : Single ion energy using different cutoff methods. [JCP, **140**, 164106 (2014), JCP, **122**, 044107 (2005)]

Integration of Newton's equation

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

$$\mathbf{F} = m\mathbf{a} - \nabla U \quad \text{Newton's Law} \quad (16)$$

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

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \delta t \mathbf{v}(t + \frac{1}{2} \delta t) \quad (17)$$

$$\mathbf{v}(t + \frac{1}{2} \delta t) = \mathbf{v}(t - \frac{1}{2} \delta t) + \delta t \mathbf{a} \quad (18)$$

velocities are updated according to,

$$\mathbf{v}(t) = \frac{1}{2} \left(\mathbf{v}(t + \frac{1}{2} \delta t) + \mathbf{v}(t - \frac{1}{2} \delta t) \right) \quad (19)$$

Constraints

Collision of two diatomic molecules

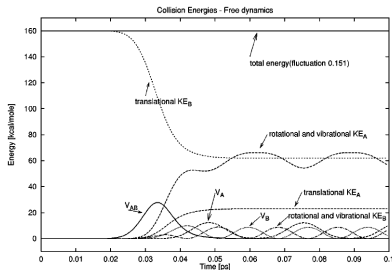


Figure : Free collision.

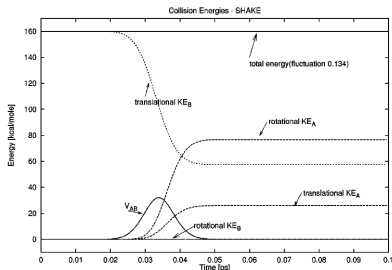


Figure : SHAKE constraint.

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

Constraints

Modern approaches to deal with constraints

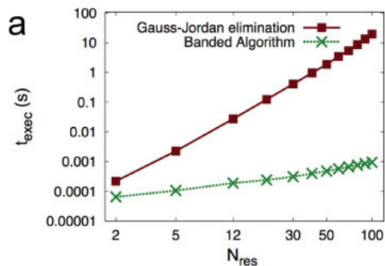


Figure : ILVES method.

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

Techniques to speedup simulations

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

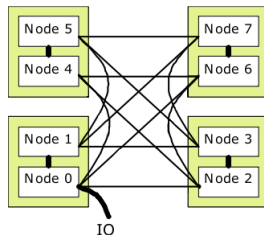


Figure : Nodes (MPI).

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

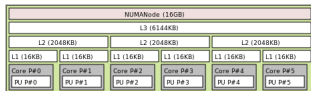


Figure : NUMA machine (OpenMP).
Computational Chemistry - MD Simulations

Ergodicity

$$\begin{aligned} \mathcal{A}_{obs} &= \langle \mathcal{A} \rangle_{\text{time}} \\ &= \langle \mathcal{A}(\Gamma(t)) \rangle_{\text{time}} \\ &= \lim_{t_{obs} \rightarrow \infty} \frac{1}{t_{obs}} \int_0^{t_{obs}} \mathcal{A}(\Gamma(t)) dt \end{aligned} \quad (20)$$



Figure : Coffee cup.

Statistical ensembles

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

$$Q_{NVE} = \frac{1}{N!} \frac{1}{h^{3N}} \int d\mathbf{r} d\mathbf{p} \delta(\mathcal{H}(\mathbf{r}, \mathbf{p}) - E) \quad (21)$$

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,

$$Q_{NVT} = \frac{1}{N!} \frac{1}{h^{3N}} \int d\mathbf{r} d\mathbf{p} \exp(-\mathcal{H}(\mathbf{r}, \mathbf{p})/k_B T) \quad (22)$$

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

Statistical ensembles

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

$$Q_{NPT} = \frac{1}{N!} \frac{1}{h^{3N}} \frac{1}{V_0} \int dV \int d\mathbf{r} d\mathbf{p} \exp(-(\mathcal{H}(\mathbf{r}, \mathbf{p}) + PV)/k_B T) \quad (23)$$

the corresponding thermodynamic potential is

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

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

$$Q_{\mu VT} = \sum_N \frac{1}{N!} \frac{1}{h^{3N}} \exp(\mu N/k_B T) \int d\mathbf{r} d\mathbf{p} \exp(-\mathcal{H}(\mathbf{r}, \mathbf{p})/k_B T) \quad (24)$$

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.

$$H = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) + \frac{p_\xi^2}{2Q} + N_f kT\xi \quad (25)$$

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,

$$V^*(\mathbf{r}) = \begin{cases} V(\mathbf{r}), & V(\mathbf{r}) \geq E, \\ V(\mathbf{r}) + \Delta V(\mathbf{r}), & V(\mathbf{r}) < E. \end{cases} \quad (26)$$

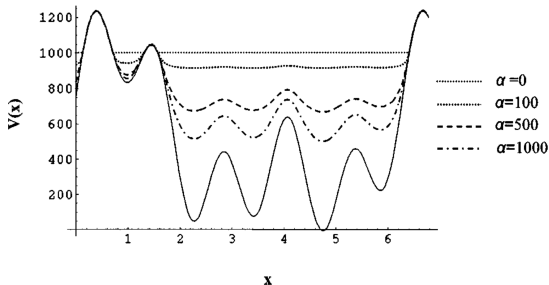


Figure : Modified potential energy surface [JCP, **120**, 11919 (2004)].

Accelerated MD simulations

the biasing term is,

$$\Delta V(\mathbf{r}) = \frac{(E - V(\mathbf{r}))^2}{\alpha + (E - V(\mathbf{r}))} \quad (27)$$

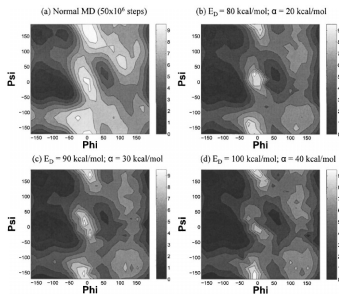


Figure : Free energy landscape of Alanine dipeptide [JCP, **120**, 11919 (2004)].

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,

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

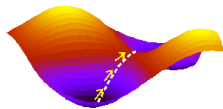


Figure : Potential energy surface.

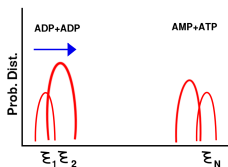


Figure : Probability histograms.

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),

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

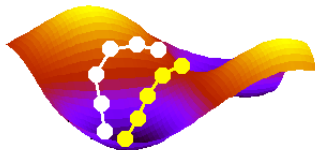


Figure : Free energy surface.

String method (SM) simulations

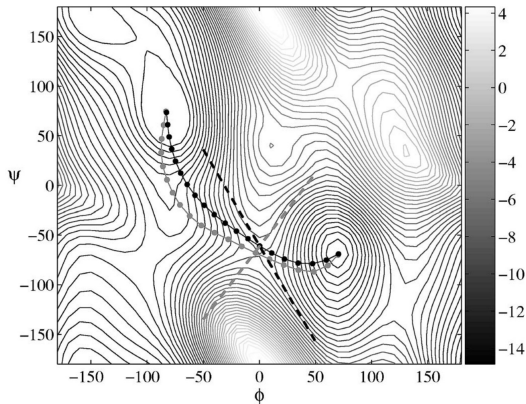


Figure : Free energy surface of Alanine dipeptide.

Coarse-grain simulations

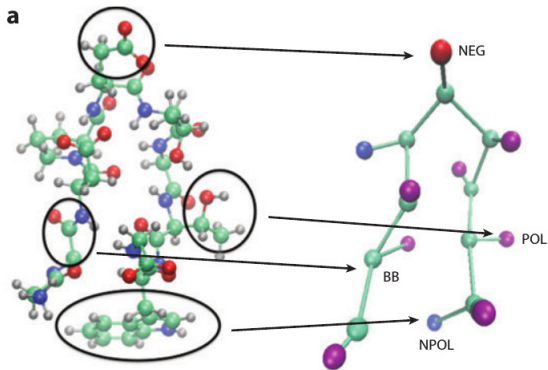


Figure : Reduction of the degrees of freedom [Annu. Rev. Biophys., **42**, 73 (2013)].

Alchemical simulations

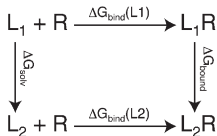


Figure : Thermodynamic cycle for binding of two protein ligands L_1 and L_2 , [JCC, 30, 1692 (2009)].

$$\Delta\Delta G_{L_i \rightarrow L_j}^{\text{bind}} = \Delta G_{L_j}^{\text{bind}} - \Delta G_{L_i}^{\text{bind}} = \Delta G_{RL_i \rightarrow RL_j}^{\text{prot}} - \Delta G_{L_i \rightarrow L_j}^{\text{solv}} \quad (28)$$

The Hamiltonian is modified according to,

$$H = T_x + (1 - \lambda)V_0 + \lambda V_1 \quad (29)$$

Alchemical simulations

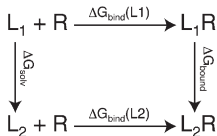


Figure : Thermodynamic cycle for binding of two protein ligands L_1 and L_2 , [JCC, 30, 1692 (2009)].

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

$$\Delta G_{\lambda=0 \rightarrow \lambda=1} = \sum_{\lambda=0}^1 -\frac{1}{\beta} \ln \langle \exp(-\beta(H_{(\lambda+\delta\lambda)} - H_{(\lambda)})) \rangle \quad (30)$$

Abisko and Intro to Kebnekeise

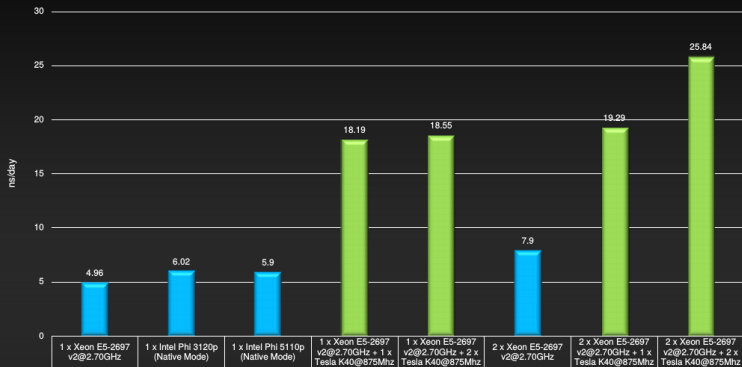
Birgitte Brydsoe

GROMACS on GPU/Phi

GROMACS 5.0: Phi vs. Kepler K40 fastest GPU!



GROMACS 5.0 RC1 (ns/day) on K40 with Boost Clocks and Intel Phi
192K Waters Benchmark (CUDA 6.0)



GROMACS

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

GROMACS files

- *.gro, *.pdb (coordinates)
- *.top (topology)
- *.tpr (binary input file)
- *.mdp (parameter file for simulation)