Basic Principles of Molecular Dynamics (MD) Theory

Dr Karina Kubiak - Ossowska
ARCHIE-WeSt, University of Strathclyde
e-mail: karina.kubiak@strath.ac.uk

Outline

1. Motivation
    a) Computer Simulations
    b) Molecular Dynamics simulations
        i. Why are Simulations needed?
        ii. Some Examples
2. Molecular Dynamics Model
3. Atom Model
4. The Potential and the Force Field
5. Energy Minimization
6. Integration Algorithms
7. Ensembles, Thermostates, Barostates
8. More advanced:
    a) Periodic Boundary Conditions (PBC)
    b) Full Electrostatic Computation (PME)
    c) Constrains
    d) Steered MD

Computer Simulations

We want to model the complex processes occurring in nature as accurately as possible. A good model has to provide a precise description of the process and to allow the prediction of the future behavior of the system (or prediction the results of the similar process) within certain bounds.

Most natural processes cannot be described using simple models. For example, the analytical solution of three-body motion does not exist.

http://en.wikipedia.org/wiki/Milky_Way

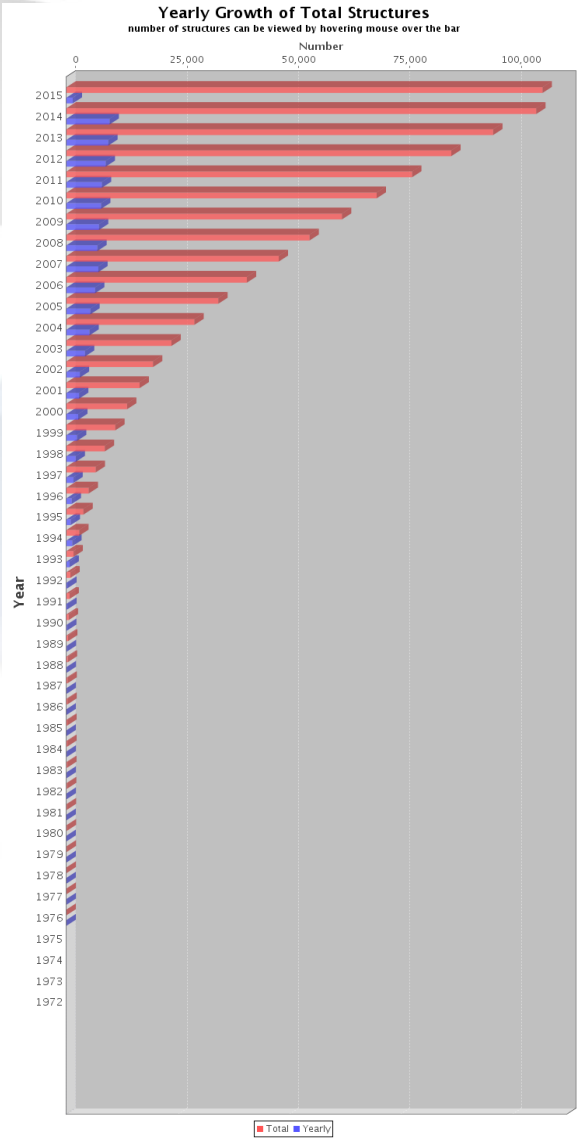
$1m^3$ of gas under normal conditions contains $2.68678x10^{25}$ atoms, 12g of carbon C_{12} contains 6.02214x1023 atoms, Milky Way consists 200 billion stars…..

Computer Simulations

Computer simulations are used for investigations of matter at the molecular level ($s=10^{-9}m$, $t=10^{-12}$ – $10^{-3}s$, $m=10^{-27}kg$) as well as studying the shape of the universe ($s=10^{23}m$, $t=10^{17}s$, $m=10^{40}kg$). Besides experimental and theoretical approaches, computer simulations are a third way in science.

Computer simulation means the mathematical prediction of physical process on a computer.

Molecular Dynamics Simulations



We want access to dynamical information extrapolated from the biomolecule structure (DNA, proteins) using a molecular modelling tool.

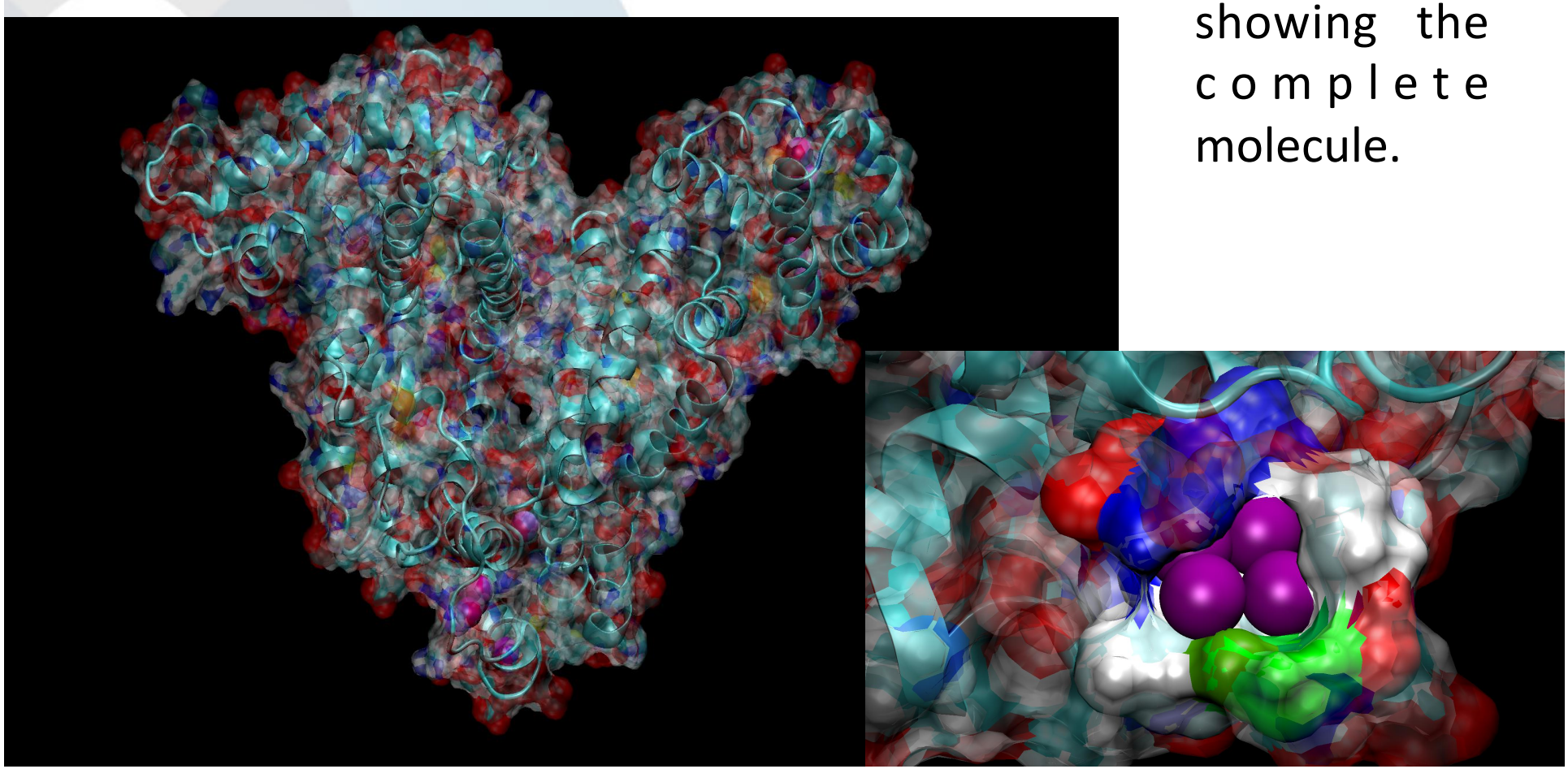
Structures deposited in the Protein Data Bank (PDB) at 25/02/2015: **106,858 including 99,287 protein structures.**

How these biomolecules behave? How they interacts with other biomolecules and materials?

http://www.rcsb.org/pdb/statistics/contentGrowthChart.do?content=total&seqid=100

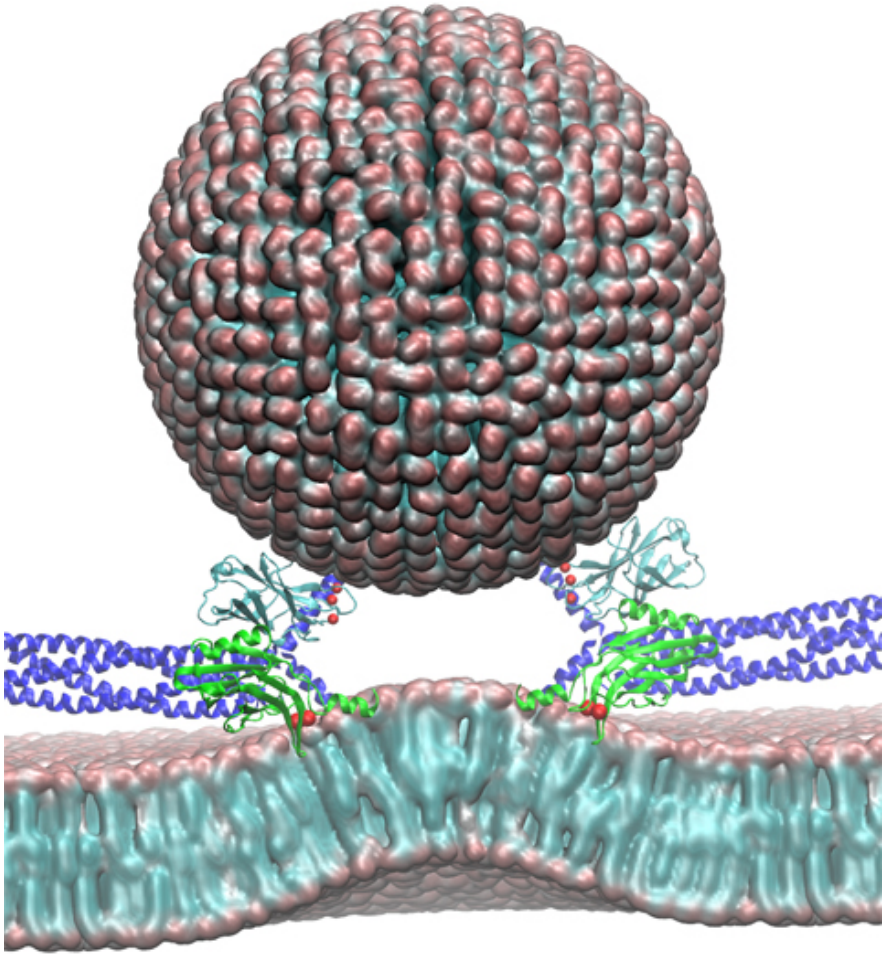
Why Simulations are Needed?

Using Molecular Dynamics we can **SEE** what biomolecules do, and what their role is in the biological system. That is the only method showing the complete molecule.



Experiment Correction

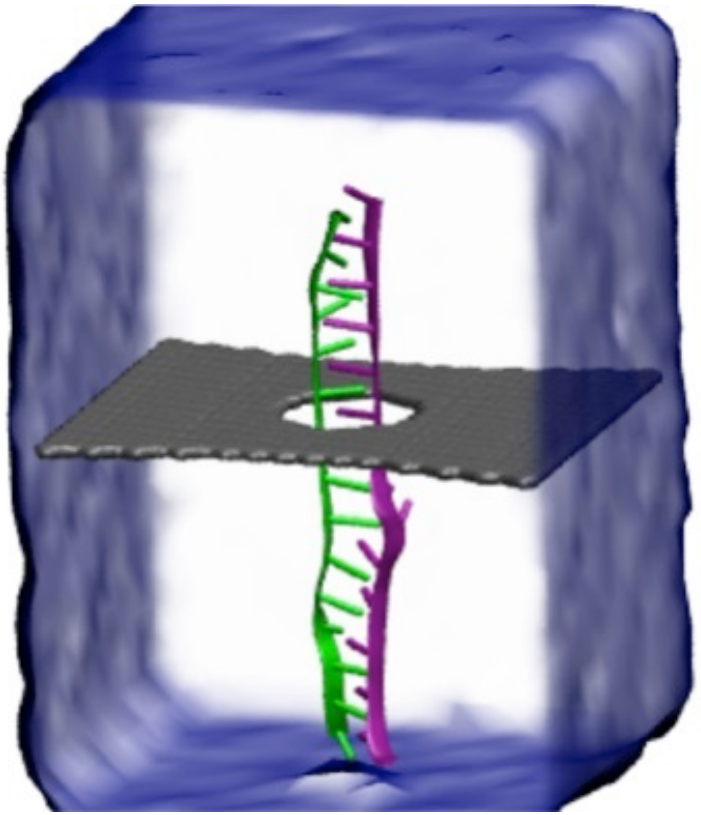
Computer simulations can correct experiments: NAMD simulations have revealed how synaptotagmin I acts (synapse neurotransmitter). Simulations suggest that the experimentally observed structure of synaptotagmin I measured in vitro in a crystal/NMR form of the protein, differs from the active in situ form of synaptotagmin I.



http://www.ks.uiuc.edu/Research/namd/spotlight/

Graphene nanopore sensor

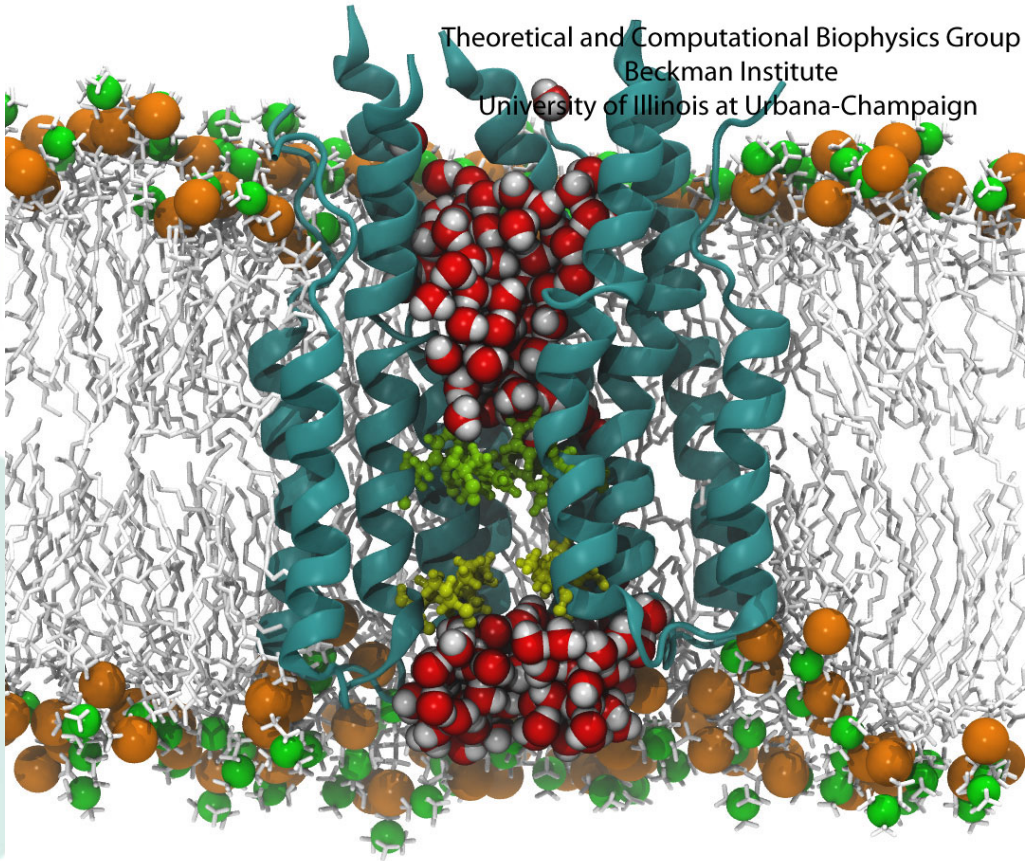
Graphene nanopores can detect translocating DNA by recording concomitant flow of charged ions through the pore. Studies suggest that sheet currents, in graphene membranes, can be used to detect conformation and sequence of a DNA molecule passing through the nanopore. This new research will guide the development of graphene-based nanosensors for DNA detection.



http://www.ks.uiuc.edu/Research/namd/spotlight/

Living cells – ion channels

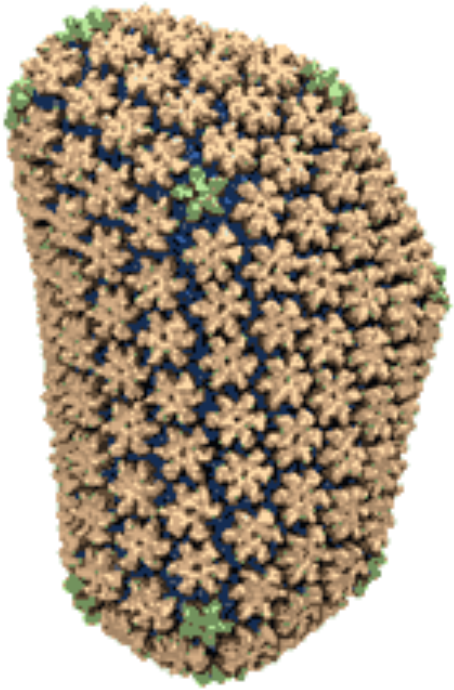
Pore proteins are very flexible, so it's very difficult to obtain a crystal structure, which is unfortunate, since they could offer attractive drug targets for new antiviral therapies. MD simulations can provide plausible 3-D models for viroporins visualization and drug design.

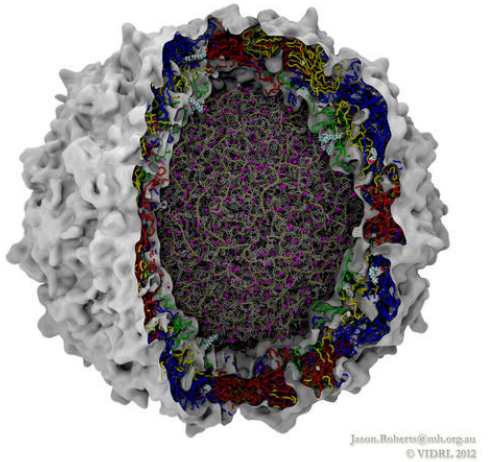


http://www.ks.uiuc.edu/Research/namd/spotlight/

Virus structure and action

HIV-1 virus is the major cause of AIDS, for which treatments need to be developed continuously (the virus quickly becomes quickly). When the virus infects a human cell it releases into the cell its capsid - a closed, stable container protecting the viral genetic material. Characterization of the HIV-1 capsid will guide the design of novel drugs.



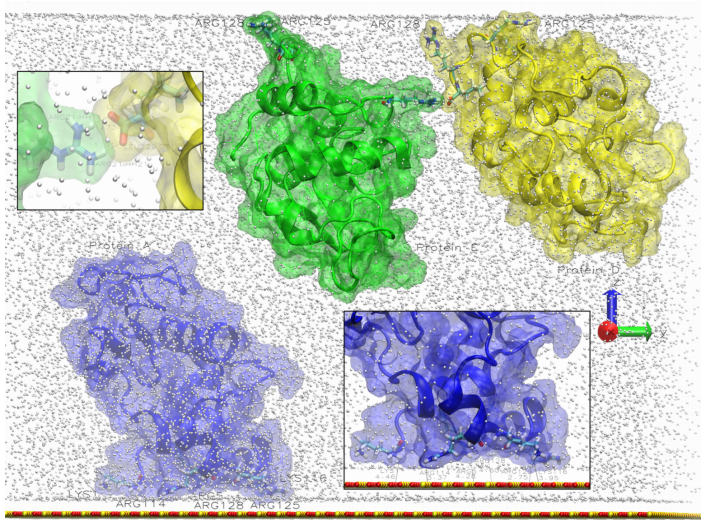


Simulations of the complete polio virus

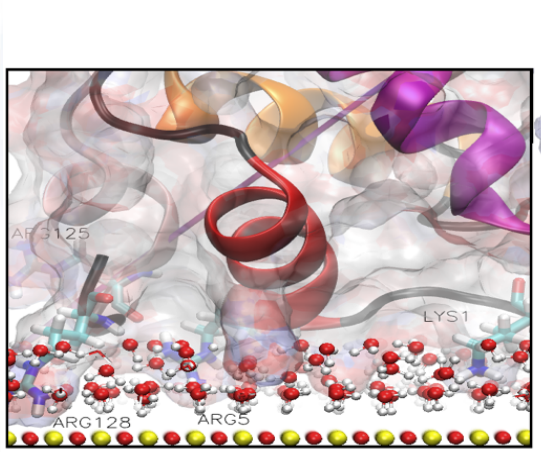
http://www.ks.uiuc.edu/Research/namd/spotlight/

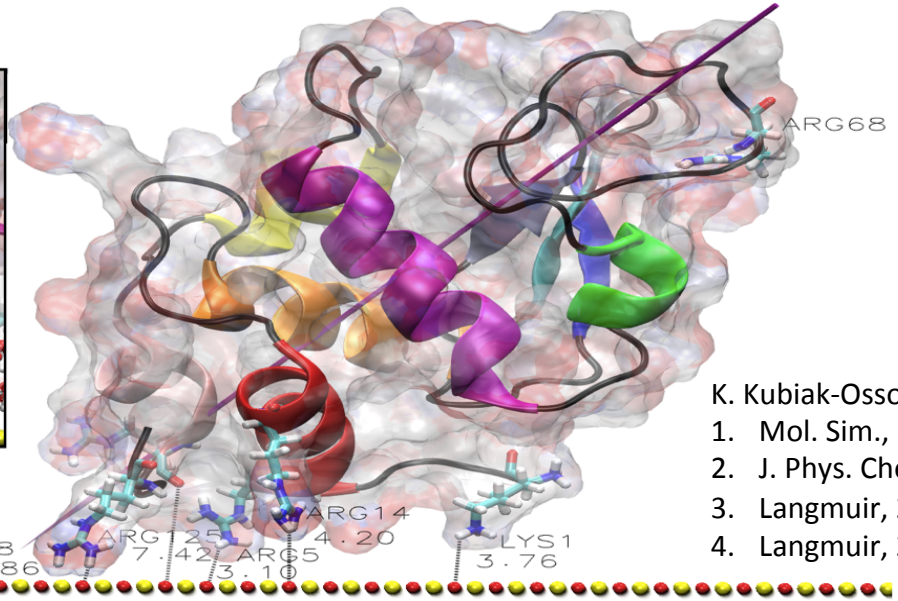
Protein adsorption

Thanks to the simulations we can see how the protein aggregates at the atomistic level and describe the adsorption and diffusion mechanisms.



K.Kubiak-Ossowska, P. Mulheran, J. Phys. Chem. B, 115, 8891-8900, 2011

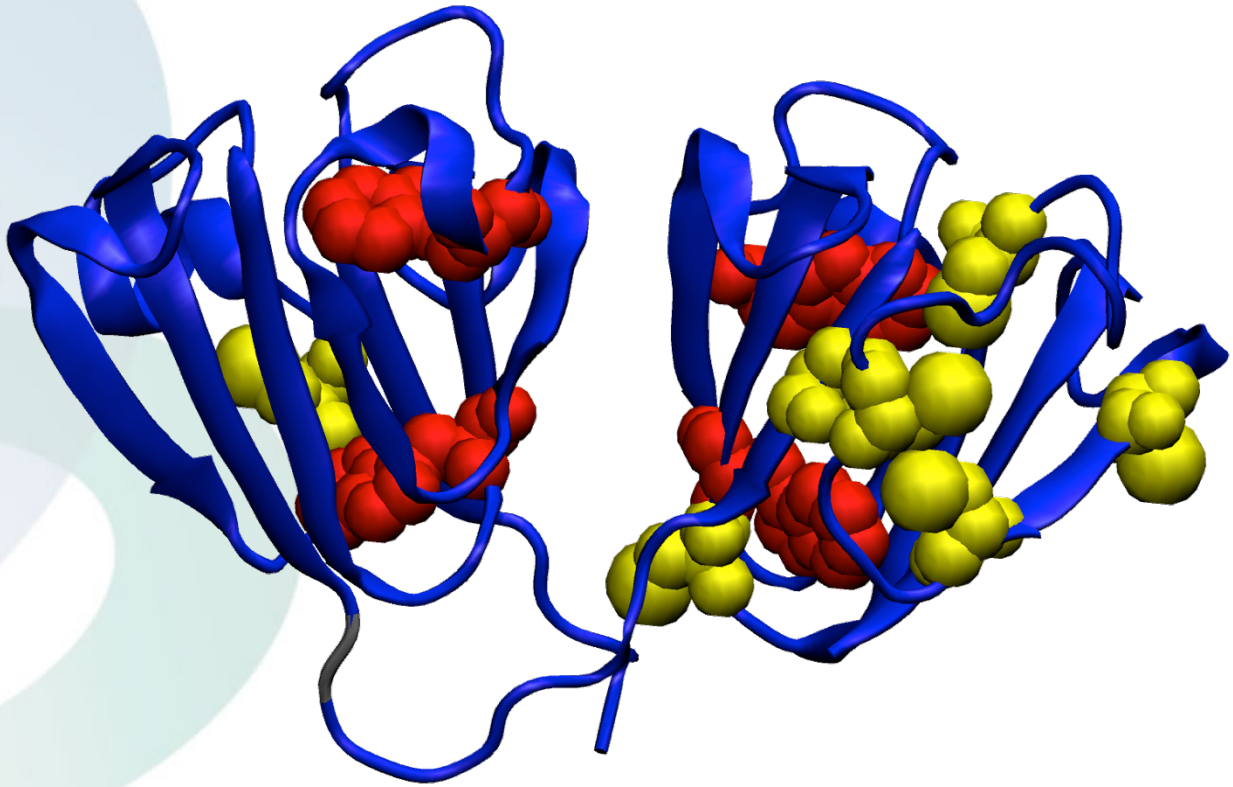




K. Kubiak-Ossowska, P. Mulheran:
1. Mol. Sim., 35, 561-566, 2009
2. J. Phys. Chem. B,113, 12189-12200, 2009
3. Langmuir, 26, 7690-7694, 2010
4. Langmuir, 26, 15954-15965, 2010

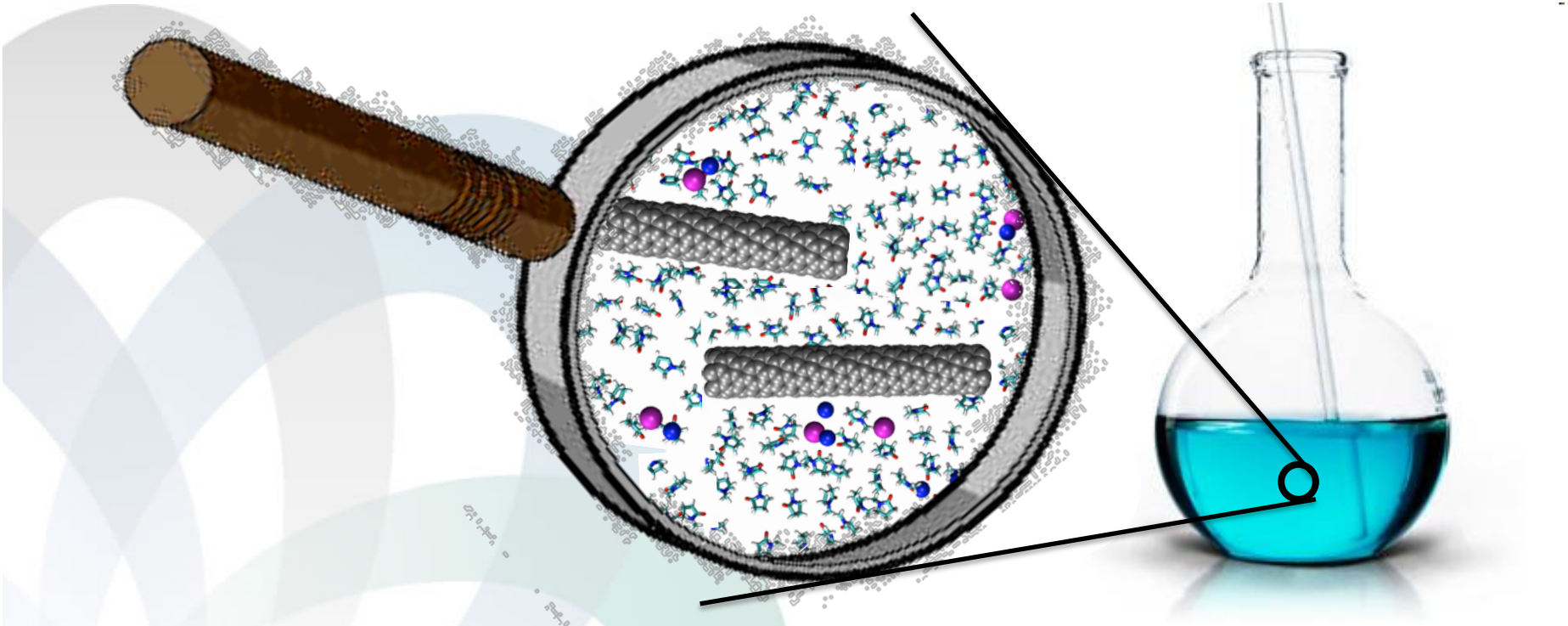
Understanding Diseases – Cataract

γ-crystalins in the human eye lens are responsible for lens transparency. Upon UV irradiation the lens becomes opaque due to crystallin aggregation. The main reason for is are TRP alterations (dye creation) that influences the protein structure and results in aggregation.



K.Kubiak, M. Kowalska, W. Nowak, J. Mol. Struct. 630, 315-325, 2003

Molecular Microscope



MD is a computational microscope enabling us to understand how things work on a detailed *molecular* level, ranging from electronic structures to long-term phasé behavior of molecules

If one understands how things work,
one can manipulate them!

But……

Molecular Dynamics is a theoretical method, to understand the meaning of results obtained we need to know the models used and be aware of the approximations made. MD is only a model, but a model which works!

To use the strength of the MD method we need to know its weakness.

MD in NAMD

MD is used to simulate molecular fluids, crystals, amorphous polymers, liquid crystals, zeolites, **nuclear acids, proteins, membranes, viruses** and many more biochemical materials.

Molecular Dynamics Model

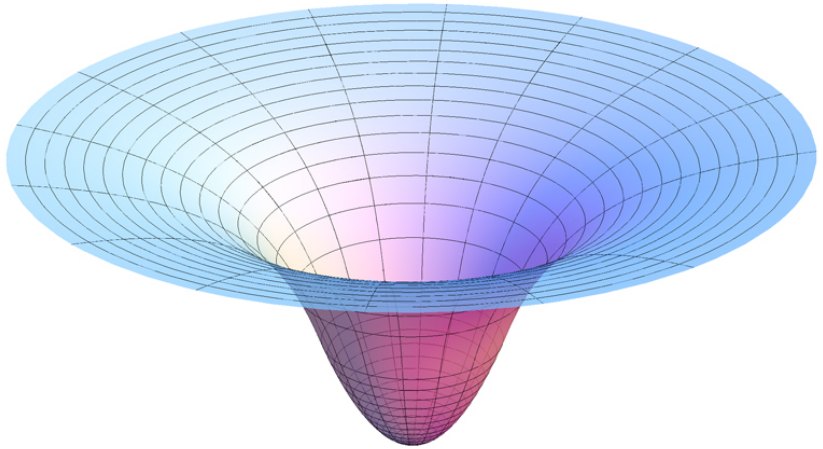
Molecular Dynamics is the computer simulation of physical movements of atoms and molecules. The trajectories of molecules and atoms are determined by **numerically solving the classical equations of motion** (Newton's equations) for a system of interacting particles, where forces between the particles and **potential energy** are defined by **molecular mechanics force fields.**

This was first accomplished in **1957 and 1959** for a system of hard spheres by Adler. Now is applied mostly in **materials science and modelling of biomolecules.**

MD and Molecular Mechanics

Molecular mechanics (MM) in MD is **used for energy minimization.** The force field is used as an optimization criterion and the **local minimum** is searched for by an appropriate algorithm (conjugate gradient, steepest descent). The main aim is **to find the lowest (or one of the lowest) energy conformation** of a molecule.

MM can be used for biomolecule optimization, design of drugs and small molecules, searching of the binding sites, protein folding kinetics …



MD Atom Model

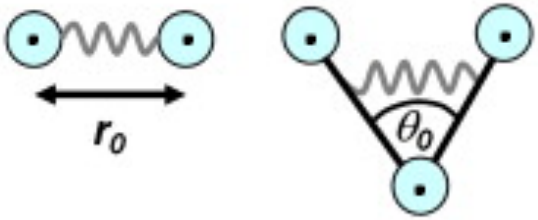
Atom – a hard ball with given radius and no directional properties and without internal degrees of freedom. The atoms are connected by virtue of the bond interaction energies.

Ball radius depends on the atom type. Sometimes is easier to treat the atoms as a point with mass and charge (equation of motion), sometimes is easier to think about the hard balls (vdW radius, calculation of the potential)

Bond – is only virtual. The bonds are not independent entities in the model and their characteristics (single/double) are introduced only through the properties of interaction with atoms.

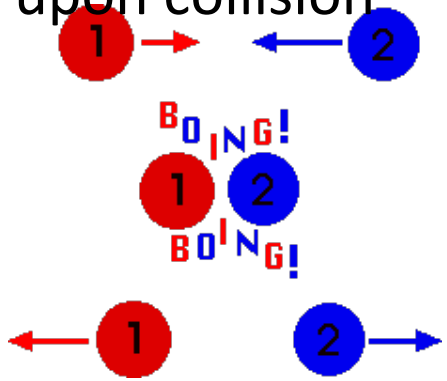
MD Atom Model

- Bonding interactions are modeled as springs (**harmonic oscillators**). All bonds and bond angles are modelled as springs.



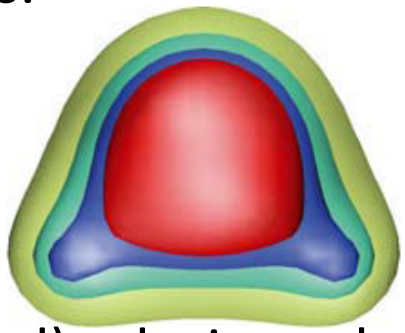
- All collisions are **perfectly elastic:**
  - i. The total kinetic energy is conserved
  - ii. The total momentum is conserved
  - iii. Particles do not change their shape upon collision

(**noble gas approximation**)



MD Atom Model

- Atomic **charges are not changed** during the trajectory
    i. Quantum effects are not included
    ii. Charge transfer is not possible
    iii. Because atoms have no directional properties the polarization effects (important in water molecules) have to be introduced in other ways.

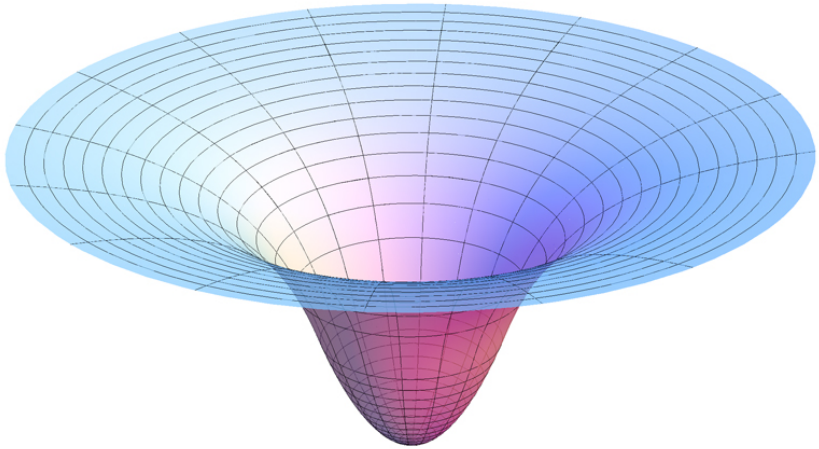


- **Bonds cannot be created** (or destroyed) during the simulation; it does not matter how far the bonded atoms are, or how close the non-bonded atoms are.

Potential Energy

Potential energy is associated with the set of forces that act on a body which **depends only on the body's position** in space. The set of forces can be considered as a **Force Field (FF)**. In other words potential energy is the energy of the object or the system due to its position. The SI unit for energy is the joule (J)

Force field is an area where the forces acts. It can be for example a gravitational force field around a uniform spherical body (picture).



Potential Energy

Protein atoms exist and move on the curved, multi-dimensional energy landscape which we want to reproduce.

Potential Energy

Evaluating the force is the most computationally demanding part of molecular dynamics.

The force is the negative gradient of a scalar potential energy function:

$$\vec{F}\left(\vec{r}\right) = -\nabla U\left(\vec{r}\right)$$

Where the potential U is a sum of bonding and nonbonding potentials:

$$U\left(\vec{r}\right) = \sum U_{bonded}\left(\vec{r}\right) + \sum U_{nonbonded}\left(\vec{r}\right)$$

U_{bonded} involves 2, 3 and 4 atom interaction while $U_{nonbonded}$ involves interactions between all pairs of atoms (ee & vdW).

Q1. Potential Energy -exercise

Let's consider a linear molecule composed from 10 atoms.

$A_1 - A_2 - A_3 - A_4 - A_5 - A_6 - A_7 - A_8 - A_9 - A_{10}$

How many pairs we have? (A_1-A_2=A_2-A_1)

Now exclude 1 – 4 atoms, how many pairs we have?

We have to do that twice: for electrostatic & L-J, so:

How many bonding interactions we have?

So how many terms in total we have?

A1. Potential Energy -exercise

Let's consider a linear molecule composed from 10 atoms.

$A_1 - A_2 - A_3 - A_4 - A_5 - A_6 - A_7 - A_8 - A_9 - A_{10}$

How many pairs we have? (A_1-A_2=A_2-A_1) 55

Now exclude 1 – 4 atoms, how many pairs we have? 21

We have to do that twice: for electrostatic & L-J, so: 42

How many bonding interactions we have? $9_{bonds}+8_{angles}+8_{dihedrals}=25$

So how many terms in total we have? 42+25=67

Q2. Potential Energy – exercise

Now let's add only 10 water molecules (30 atoms).

How many $U_{nonbond}$ water – water interactions do we have?

Do that twice, so:

How many nonbonding interactions between water and our molecule do we have?

How many water U_{bond} terms do we have?

How many terms for 10 atoms molecule and 10 water system do we have? (40 atoms)

U_{bond}=

Total=

$U_{nonbond}$=

A2. Potential Energy – exercise

Now let's add only 10 water molecules (30 atoms).

How many $U_{nonbond}$ water – water interactions do we have? 405

Do that twice, so: 810

How many nonbonding interactions between water and our molecule do we have? 2x10x10=600

How many water U_{bond} terms do we have? 3x10=30

How many terms for 10 atoms molecule and 10 water system do we have? (40 atoms)

U_{bond}= $25_{molecule}+30_{water}=55$

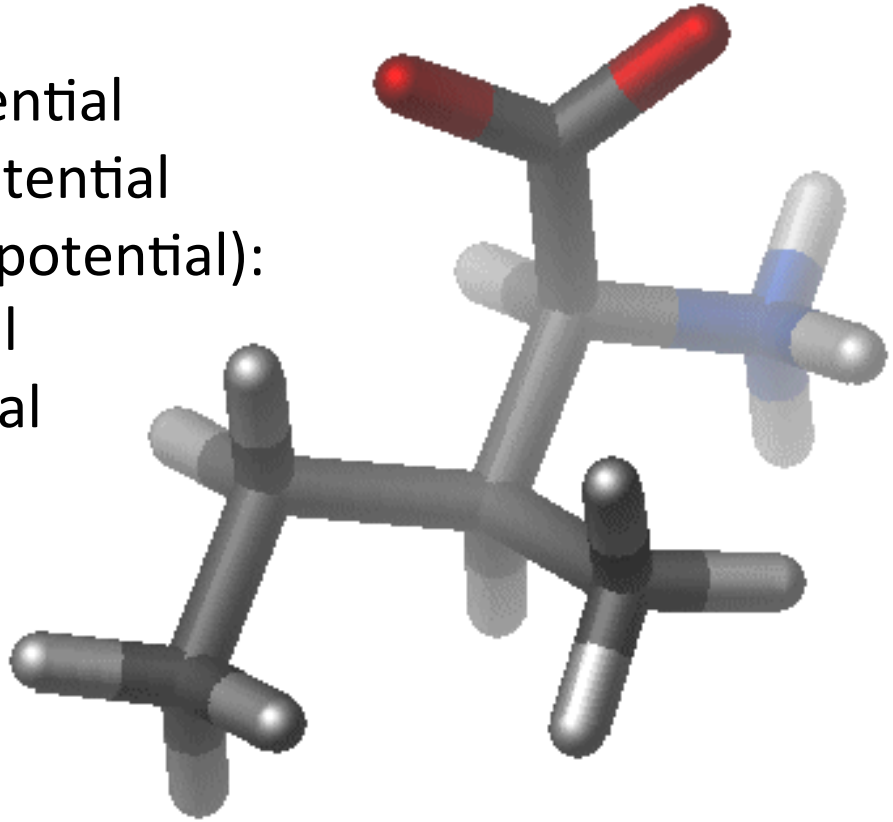
Total= 1507

$U_{nonbond}$= $42_{molecule}+810_{water\text{-}water}+600_{water\text{-}molecule}=1452$

U_{bonded}

The bonded potential terms involve 2, 3, and 4 body interactions of covalently bonded atoms. They include:

1. Two body spring bond potential
2. Three body spring angle potential
3. Four body (torsional angle potential):
   a. Dihedral angle potential
   b. Improper angle potential



http://www.ch.embnet.org/MD_tutorial/pages/MD.Part2.html

$U_{bonding}$

We use simple harmonic oscillator approximation (neither driven nor damped):

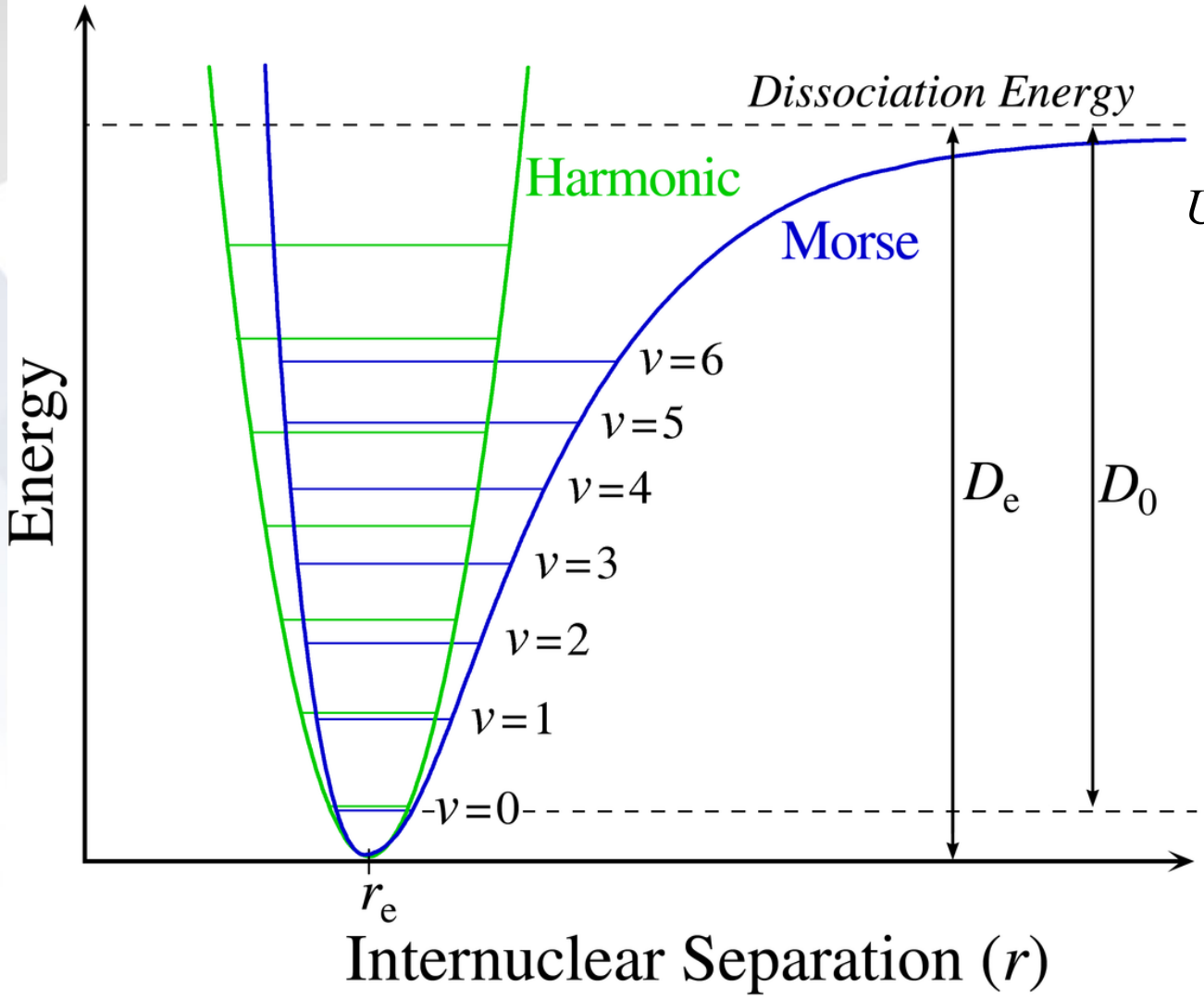
$$F = -kx \qquad U = \frac{1}{2}kx^2$$

More realistic would be Morse potential, which explicitly includes bond breaking:

$$U(r) = D_e\left(1 - e^{-\alpha(r-r_e)^2}\right) \qquad \alpha = \sqrt{\frac{k_e}{2D_e}}$$

But it is much more difficult to calculate……

Harmonic vs Morse potential



$$U(r) = D_e\left(1 - e^{-\alpha(r-r_e)^2}\right)$$

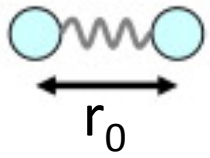
$$\alpha = \sqrt{\frac{K_e}{2D_e}}$$

D_e – the well depth
α - the well width
K_e - the force constant at the minimum of the well

Bond Potential

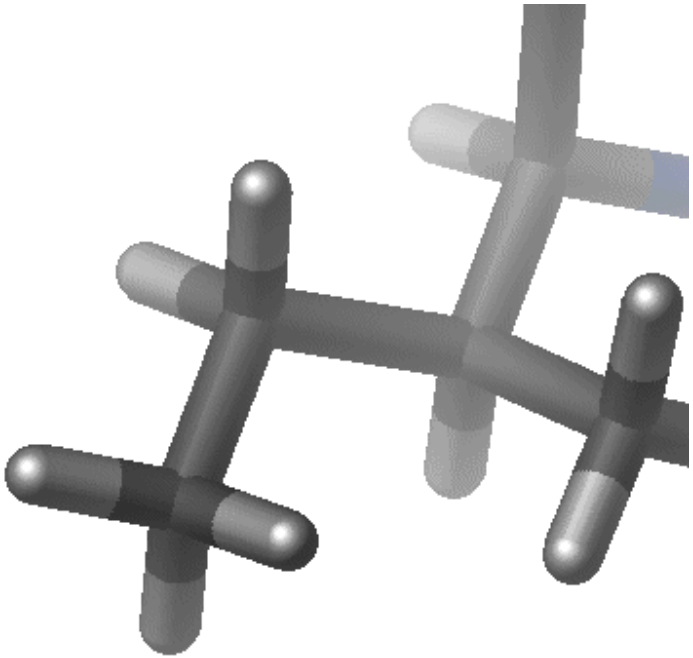
The 2-body spring bond potential describes the harmonic vibrational motion between an *(i, j)* pair of covalently bonded atoms:

$$U_{bonds}\left(\vec{R}\right) = \frac{1}{2}\sum_{bonds} K_b\left(r_{i,j} - r_0\right)^2$$



Where $r_{i,j} = \left\|\vec{r_j} - \vec{r_i}\right\|$ is the distance

Between the atoms, r_0 is the equilibrium distance and K_b is the spring constant.



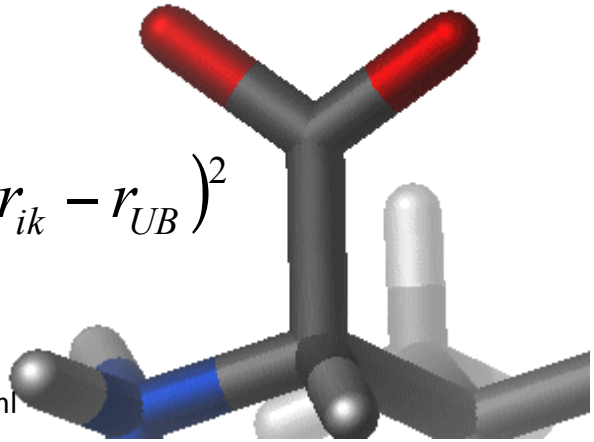
http://www.ks.uiuc.edu/Research/namd/2.7/ug/

http://www.ch.embnet.org/MD_tutorial/pages/MD.Part2.html

Angle Potential

The 3-body angular bond potential describes the angular vibrational motion occurring between an *(i,j,k)* triple of covalently bonded atoms:

$$U_{angles}\left(\vec{R}\right)=\frac{1}{2}\sum_{angles}K_{\theta}\left(\theta-\theta_{0}\right)^{2}+K_{UB}\left(r_{ik}-r_{UB}\right)^{2}$$



http://www.ch.embnet.org/MD_tutorial/pages/MD.Part2.html

Where θ in the angle (in radians) between vectors:

$$\vec{r}_{ij}=\vec{r}_{j}-\vec{r}_{i}\quad and\quad \vec{r}_{kj}=\vec{r}_{j}-\vec{r}_{k}$$

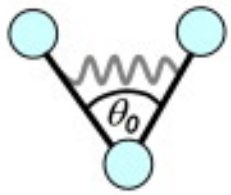
θ_0 is the equilibrium angle and K_θ is the spring constant.

http://www.ks.uiuc.edu/Research/namd/2.7/ug/

Angle Potential

The second term is the Urey-Bradley term used to describe a noncovalent spring between the outer *i* and *k* atoms, active when constant $K_{UB} \neq 0$, where $r_{ik} = \left\| \vec{r_k} - \vec{r_i} \right\|$ gives the distance between the pair of atoms and r_{UB} is the equilibrium distance.

$$U_{angles}\left(\vec{R}\right) = \frac{1}{2}\sum_{angles} K_\theta \left(\theta - \theta_0\right)^2 + K_{UB}\left(r_{ik} - r_{UB}\right)^2$$



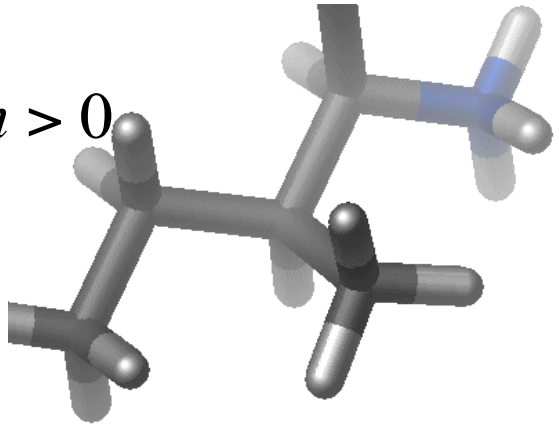
http://www.ks.uiuc.edu/Research/namd/2.7/ug/

Torsional Angle Potential

The 4-body torsion angle (also known as dihedral angle) potential describes the angular spring between the planes formed by the first three and last three atoms of a consecutively bonded *(i,j,k,l)* quadruple of atoms

$$U_{tors}\left(\vec{R}\right) = \sum_{dih} K_{\psi}\left(1+\cos\left(n\psi+\phi\right)\right) \quad if \quad n>0$$

$$= \sum_{imp} K_{\psi}\left(\psi-\phi\right)^2 \quad if \quad n=0$$



http://www.ch.embnet.org/MD_tutorial/pages/MD.Part2.html

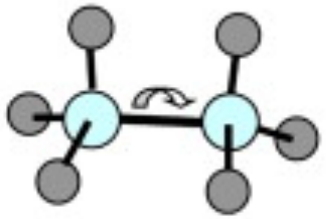
where Ψ is the angle in radians between the *(i,j,k)*-plane and the *(j,k,l)*-plane. The integer constant *n* is nonnegative and indicates the periodicity. For *n>0*, ϕ is the phase shift angle and K_{ψ} is the multiplicative constant. For n=0, ϕ acts as an equilibrium angle (it's units change to potential/rad^2). http://www.ks.uiuc.edu/Research/namd/2.7/ug/

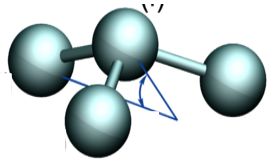
Torsional Angle Potential

A given *(i,j,k,l)*-quadruple of atoms might contribute multiple terms to the potential, each with its own parameterization. The use of multiple terms for a torsion angle allows for complex angular variation of the potential which is effectively a truncated Fourier series.

$$U_{tors}\left(\vec{R}\right) = \sum_{dih} K_{\psi}\left(1+\cos\left(n\psi+\phi\right)\right) \quad if \quad n>0$$

$$= \sum_{imp} K_{\psi}\left(\psi-\phi\right)^2 \quad if \quad n=0$$





http://www.ks.uiuc.edu/Research/namd/2.7/ug/

U_{bonded} - summary

So the bonding potential is given by:

$$U = U_{bonds} + U_{angles} + U_{tors}$$

$$U = \frac{1}{2}\sum_{bonds} K_b\left(r_{ij} - r_0\right)^2 + \frac{1}{2}\sum_{angles} K_\theta\left(\theta - \theta_0\right)^2 + K_{UB}\left(r_{ik} - r_{UB}\right)^2$$

$$+ \sum_{dih} K_\psi\left(1 + \cos\left(n\psi + \phi\right)\right) + \sum_{imp} K_\psi\left(\psi - \phi\right)^2$$

http://www.ks.uiuc.edu/Research/namd/2.7/ug/

$U_{nonbonded}$

The nonbonded potential terms involve interactions between all *(i,j)*-pairs of atoms, usually excluding pairs of atoms already involved in a bonded term. Only the van der Waals and electrostatic interactions are considered.

$$U_{nonbonded} = U_{vdW} + U_{el}$$

Even using fast evaluation methods, the cost of computing the nonbonded potentials dominates the work required for each time step of an MD simulation. (see our example above).

http://www.ks.uiuc.edu/Research/namd/2.7/ug/

Van der Waals Interactions

The **van der Waals force** is the sum of the attractive or repulsive forces between atoms other than those due to covalent bonds or the electrostatic interaction of ions or charged atoms.

The vdW force includes:

- The force between two permanent dipoles (Keesom force)
- The force between a permanent dipole and a corresponding induced dipole (Debye force)
- The force between two instantaneously induced dipoles (London dispersion force)

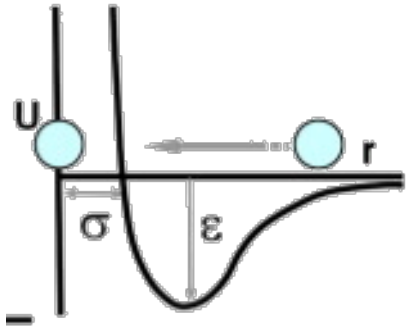
Van der Waals forces are relatively weak compared to covalent bonds or electrostatic interactions, but play a fundamental role chemistry, biology, nanotechnology, surface science.

http://en.wikipedia.org/wiki/Van_der_Waals_force

Van der Waals Interactions

The Van der Waals Force is described by the Lenard-Jones Potential:

$$U_{vdW} = \left(-\varepsilon_{\min}\right)\left[\left(\frac{R_{\min}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min}}{r_{ij}}\right)^{6}\right]$$



repulsion

attraction

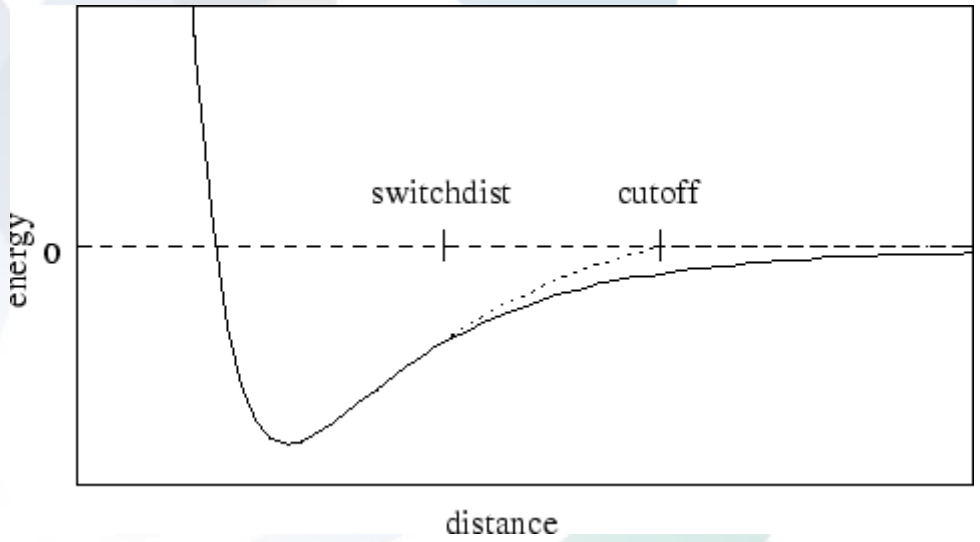
Where ε_{min} is the minimal potential observed (well depth).

The Lennard-Jones potential approaches 0 rapidly as r_{ij} increases, so it is usually truncated (smoothly shifted) to 0 past a cutoff radius.

http://www.ks.uiuc.edu/Research/namd/2.7/ug/

$U_{nonbonded}$ in Practice

In NAMD `cutoff` is used together with `switchdist`, which means that the potential is smoothly extrapolated to zero between the values `switchdist` and `cutoff`.



http://www.ks.uiuc.edu/Research/namd/2.7/ug/

Force Field

The exact formulation of the potential energy formula and the choice of parameters defines the force field.

Each spring constant (K_B, K_θ, K_{UB}, K_ψ) and each equilibrium value (r_0, θ_0, r_{UB}, ϕ) has to be defined for each pair, triple and quadruple of atoms. Moreover each atom needs to be accompanied with the charge (q)and LJ parameters (R_{min}, ε_{min}).

The determination of these parameters requires significant work undertaken as a combination of empirical and quantum mechanical calculations. Then the force field has to be tested to check if reproduces correctly the structural, dynamical and thermodynamical properties of molecules that have been well-characterized experimentally.

Force Field

Commonly used Force Fields in Molecular Dynamics (including the energy minimization are: **AMBER, CHARMM, GROMOS, OLPS, CVFF**. There are also polarizable versions.

Force fields for coarse grained simulations:

VAMM – knowledge based force field for CG simulations

MARTINI – for coarse grained MD simulations of lipids and proteins

Energy minimization

Energy minimization methods are used to **compute the equilibrium configuration of molecules**. Stable states of molecules correspond to global and local minima on their potential energy surface. Starting from non-equilibrium geometry the mathematical procedure of optimization moves atoms to find the **lowest energy configuration**.

Energy minimization **does not include the temperature**. From physical point of view the final state of the system corresponds to the configuration of atoms in zero temperature.

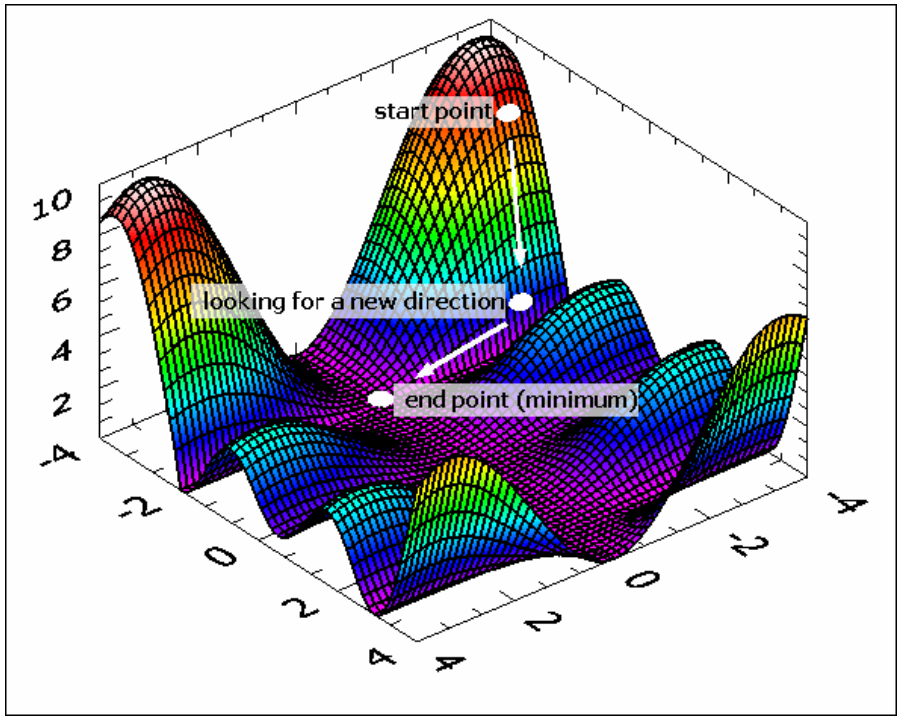
Potential energy surface – multidimensional surface (hypersurface) with atomic positions as variables. During the minimization procedure we change atoms coordinates to find the global (ideally) minima.

Energy minimization

In NAMD The default minimizer uses a sophisticated **conjugate gradient** and line search algorithm. The method of conjugate gradients is used to select successive search directions (starting with the initial gradient) which eliminate repeated minimization along the same directions.

Once the minimum in the given direction is found the algorithm looks for the next, search direction.

It is simple and … just works!



MD trajectory

MD trajectory is a series of atoms positions (snapshots) in time.
To obtain the trajectory we need to :

1. Build the **molecule** – find the (*x,y,z*) position of all atoms
2. Chose the **force field** (or create it)
3. Find the optimal structure (**energy minimization**)
4. **Calculate the trajectory = Solve the Newton's equations of motions** of all atoms in the force field at each time moment (timestep)
5. **Analyze** the trajectory obtained, make or verify hypothesis and … write the NATURE paper ☺

Newton's equations

The force field defines the potential energy $U(\vec{R})$ where $\vec{R} = \{\vec{r}_1, \vec{r}_2, \vec{r}_3, \ldots, \vec{r}_N\}$ is a set of separate atom vectors $\vec{r}_1 = (x, y, z)$ and N is a number of atoms in the system.

We know that the force acting on atom i is:

$$\vec{F}(\vec{r}_i) = -\nabla U(\vec{r}_i)$$

and simultaneously from Newton's second law:

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

and we want to find $\vec{r}_i(t)$ (solve the equation on motion) for atoms i=1, 2,...,N

Integration Algorithms

1. Finite Difference Methods
2. The Verlet Algorithm
3. **Velocity Verlet Algorithm (leap-frog)**
4. ….. Many others…….

Finite Difference Method

Given the position of the atoms, velocities and other dynamical information at time t we attempt to obtain the positions, velocities, etc. at the later time t+δt. The equations are solved on a step-by-step basis using the Taylor expansion:

$$r^p(t+\delta t) = r(t) + \delta t \bullet v(t) + \frac{1}{2}\delta t^2 \bullet a(t) + \frac{1}{6}\delta t^3 \bullet b(t) + \ldots$$

$$v^p(t+\delta t) = v(t) + \delta t \bullet a(t) + \frac{1}{2}\delta t^2 \bullet b(t) + \ldots$$

$$a^p(t+\delta t) = a(t) + \delta t \bullet b(t) + \ldots$$

$$b^p(t+\delta t) = b(t) + \ldots$$

Finite Difference Method

The algorithm would be:

1. Predict the positions, velocities, accelerations at the time t +δt using current values of these quantities
2. Evaluate the forces and then the accelerations from the new positions
3. Correct the predicted positions, velocities, accelerations using the new accelerations
4. Calculate variables (for example energy) and go to step 1

The Velocity Verlet Algorithm

The Velocity Verlet algorithm is a **numerical method used to integrate Newton's equations of motion**.

1. Fast and accurate
2. Linear and angular momentum preserved
3. Time reversible
4. Requires only one force evaluation for each time step
5. The algorithm is symplectic (long-time stable)

The Velocity Verlet algorithm generates a sequence of "snapshots" for the particle coordinates and velocities at all intermediate times Δt

The Velocity Verlet Algorithm

The position of the atom is updated every Δt step

$$\vec{r}_i(t+\Delta t) = \vec{r}_i(t) + \vec{v}_i(t)\Delta t + \frac{1}{2}\vec{a}_i(t)\Delta t^2$$

while the velocity is updated every $\left(\Delta t + {}^{\Delta t}/_{2}\right)$ step:

$$\vec{v}_i\left(t+\frac{\Delta t}{2}\right) = \vec{v}_i(t) + \frac{1}{2}\vec{a}_i(t)\Delta t$$

The acceleration in the next step $(t+\Delta t)$:

$$\vec{a}_i(t+\Delta t) = -\left(\frac{1}{m_i}\right)\nabla U\left(\vec{r}_i(t+\Delta t)\right)$$

The velocity in the next step $(t+\Delta t)$:

$$\vec{v}_i(t+\Delta t) = \vec{v}_i\left(t+\frac{\Delta t}{2}\right) + \frac{1}{2}\vec{a}_i(t+\Delta t)\Delta t$$

Trajectory calculations

For atom i=1:

1. Calculate the initial acceleration $\vec{a}_1(t)$ from the force field

$$\vec{a}_1(t) = -\left(\frac{1}{m_1}\right)\nabla U(\vec{r}_1(t))$$

2. Guess the initial velocity $\vec{v}_1(t)$ (Gaussian distribution with a temperature dependent variance)
3. Calculate the position at next time interval $(t + \Delta t)$

$$\vec{r}_1(t + \Delta t) = \vec{r}_1(t) + \vec{v}_1(t)\Delta t + \frac{1}{2}\vec{a}_1(t)\Delta t^2$$

4. Repeat 1-3 for atoms i=2,…,N
5. Calculate the force field at the next time interval $U\left(\vec{R}(t + \Delta t)\right)$ $\vec{R} = \{\vec{r}_1, \vec{r}_2, \vec{r}_3, \ldots, \vec{r}_N\}$ using new positions of all atoms i=1,…,N.

Trajectory calculations cd

For atom i=1:

6. Calculate the acceleration $\vec{a}_1(t+\Delta t)$ at the next time interval

$$\vec{a}_1(t+\Delta t) = -\left(\frac{1}{m_1}\right)\nabla U\left(\vec{r}_1(t+\Delta t)\right)$$

7. Calculate the velocity at the next half- time interval

$$\vec{v}_i\left(t+\frac{\Delta t}{2}\right) = \vec{v}_i(t) + \frac{1}{2}\vec{a}_i(t)\Delta t$$

8. Calculate the velocity $\vec{v}_1(t+\Delta t)$ at the next time interval

$$\vec{v}_i(t+\Delta t) = \vec{v}_i\left(t+\frac{\Delta t}{2}\right) + \frac{1}{2}\vec{a}_i(t+\Delta t)\Delta t$$

9. Repeat for atoms i=2,…,N
10. Repeat 3-9
11. At given intervals rescale velocities to given temperature

MD trajectory

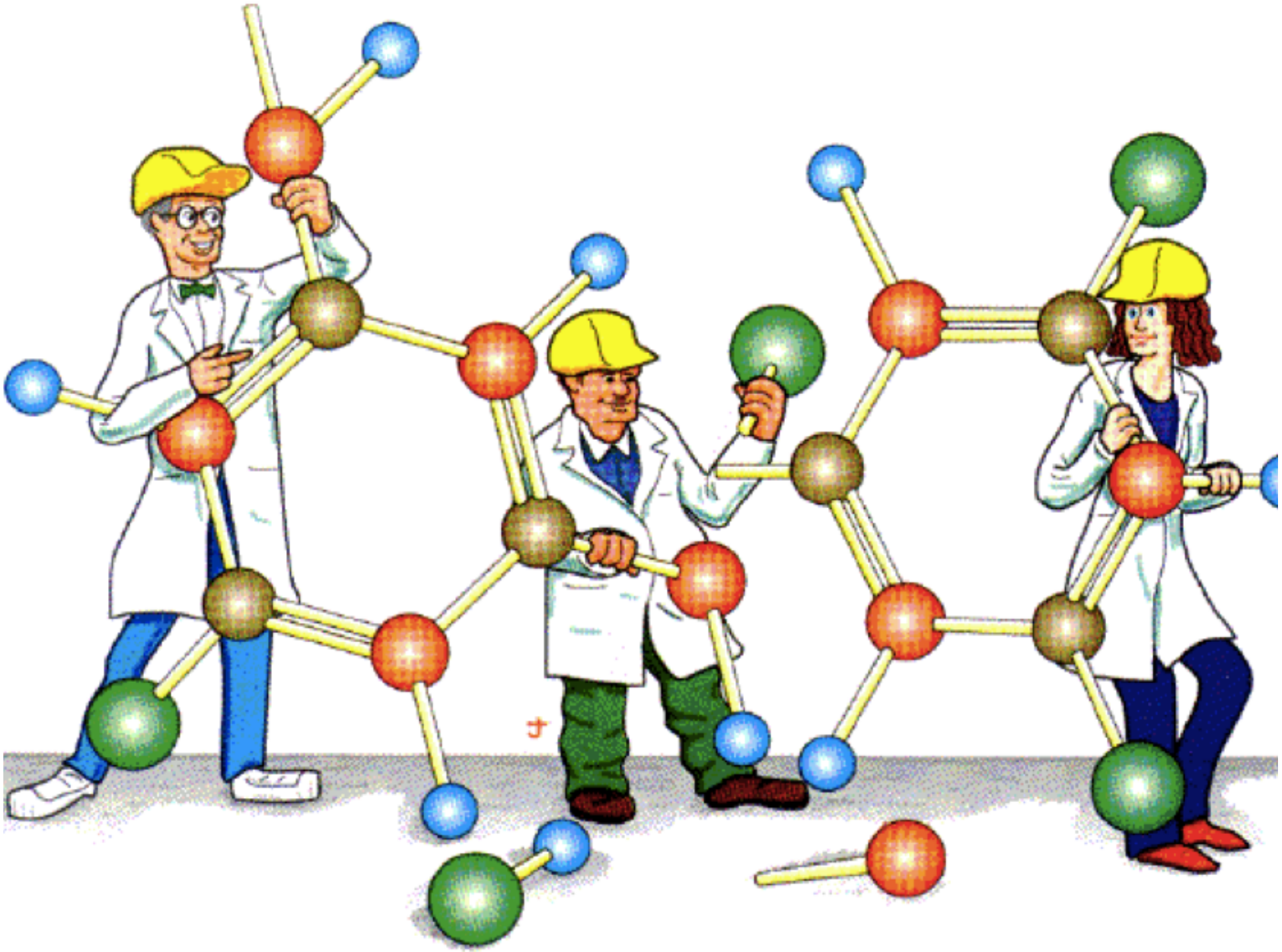
A MD trajectory is a series of atom positions (snapshots) in time.
To obtain a trajectory we need to :

1. Build the **molecule** – find the (*x,y,z*) position of all atoms
2. Chose the **force field** (or create it)
3. Find the optimal structure (**energy minimization**)
4. **Calculate the trajectory = Solve the Newton's equations of motions** of all atoms in the force field at each time moment (timestep)
5. **Analyze** the trajectory obtained, make or verify hypothesis and … write the NATURE paper ☺

To run MD we have to:

1. Build the **molecule** – find the (*x,y,z*) position of all atoms
2. Create (or apply) the existing **force field** (choose the software package)
3. Find the optimal structure (**energy minimization**)
4. **Solve Newton's equations of motions** for all atoms in the force field at each time moment (timestep) -> the trajectory
5. **Analyze the trajectory** obtained, make or verify hypothesis and … write the paper ☺

Except for 5, we already have all the information needed. Let's summarize



Build the protein?

Building the protein is not very good idea - there are too many atoms each with (*x,y,z*) positions. However, the 3D protein structure is crucial for its activity…

How to put the protein into the computer?

Protein Data Bank (pdb). Visit the page http://www.rcsb.org/pdb/home/home.do and download the structure!

Protein in the computer

Protein structures are solved by:

1. **X-ray diffraction**. Experimentalist has to make the protein crystal first (difficult), **H-hydrogen atoms are not visible** within this method (so we have to add them)
2. **NMR.** Works well only for small molecules, the entry includes several possible conformations – we have to chose one of them

Sometimes there are numerous entries for the same protein. **How do we choose the best structure?** Check the deposition date, the resolution, experimental conditions, presence of any additives, if there are missing atoms or residues. It is easy to add the missing atoms - adding missing residues is more complicated.

Simulation Run

1. Choose the molecule from the Protein Data Bank
2. Choose the force field (choose the software)
3. **Energy minimization:** Minimize the potential energy (find the optimal conformation) -> calculate $U\left(\vec{R}(t=0)\right)$ which will be used to calculate initial accelerations in the trajectory calculations.
4. **Calculate the trajectory** under chosen conditions (temperature T, pressure P, water). Temperature and/or pressure control means that …. our equations have to be modified ☹ because in the Verlet algorithm we don't have the control of T or P.

Ensembles

In classical MD, simulations are done with a **microcanonical ensemble** (N,V,E are constant) where energy is controlled, rather than temperature. Experiments are (usually) done in the **canonical ensemble** (N,V,T are constant). It means we need to introduce the method to control the temperature – **the thermostat**.

For (N,P,T) ensemble we also need **the barostat** – the mechanism to keep the pressure constant.

Ensembles

Microcanonical ensemble (NVE) – number of atoms (N), the system volume (V) and the total energy (E) are conserved.
Canonical ensemble (NVT) - number of atoms (N), the system volume (V) and the temperature (T) are conserved. Energy is exchanged with a thermostat.
Isothermal-isobaric ensemble (NPT) - number of atoms (N), the system pressure (P) and the temperature (T) are conserved. A thermostat and a barostat are needed.

Based on positions and velocities we have to calculate statistical quantities such as temperature.

Temperature

Temperature is a **statistical quantity** which has to be expressed as a function of position and momenta of all particles in the system. For a system containing large enough number of atoms the **temperature can be estimated from the kinetic energy**:

$$\left\langle \frac{1}{2} m v^2 \right\rangle = \frac{1}{2} N_f K_B T$$

where N_f is the number of degreed of freedom in the system with N particles (atoms) and K_B is Boltzman constant =1.38x10^{-23} $\frac{m^2 kg}{s^2 K}$

Temperature depends on time because positions and velocities depend on time

$$T(t) = \sum_{i=1}^{N} \frac{m_i v_i^2(t)}{K_B N_f}$$

Ensembles

A fundamental requirement for an integrator is to generate the correct ensemble distribution for the specified temperature and pressure in an appropriate way to allow the interpretation of the computed trajectory in a conventional way. For this reason the system needs to be coupled (in deterministic or stochastic way). Typically NAMD uses the stochastic coupling approach (Langevin dynamics).

Thermostat

Thermostat is the algorithm which modifies the classical Newtonian MD.

Berendsen thermostat – the most straight forward method, not very accurate

Nosé–Hoover thermostat – efficient and accurate method for constant temperature simulations

Langevin dynamics – stochastic method commonly used in NAMD

Langevin Dynamics

Newton equations of motions are modified by adding the dissipative and the fluctuating force.

The Langevin equation:

$$m\dot{v} = F(r) - \gamma v + \sqrt{\frac{2\gamma K_B T}{m}} R(t)$$

Dissipative force

Fluctuating force

m – mass, F – force, r – position, γ – friction coefficient, K_B – Boltzmann constant, T – temperature, $R(t)$ – univariate Gaussian random process, $\dot{r} = v$ and $\dot{v} = a$

Phillips et al, J. Comput. Chem, 26, 2005

Langevin Dynamics

To integrate the Langevin equation NAMD uses Brunger-Brooks-Karplus method (BBK), which is an extension of the Verlet algorithm:

$$\vec{r}_i(t+\Delta t) = \vec{r}_i(t) + \frac{1-\frac{\gamma\Delta t}{2}}{1+\frac{\gamma\Delta t}{2}}\left\{\vec{r}_i(t) - \vec{r}_i(t-\Delta t)\right\} +$$

$$\frac{1}{1+\frac{\gamma\Delta t}{2}}\Delta t^2\left[\frac{F(\vec{r}_i(t))}{m} + \sqrt{\frac{2\gamma K_B T}{\Delta m}}Z(t)\right]$$

Phillips et al, J. Comput. Chem, 26, 2005

Barostat

Barostat is the algorithm which modifies the classical Newtonian MD in order to generate the statistical quantity such as the pressure. For NPT ensemble the barostat has to be accompanied with the thermostat.

For simulations of the NPT ensemble NAMD uses a **Nosé-Hoover method combined with Langevin dynamics** (piston method) to control fluctuations in the barostat.

Further reading:

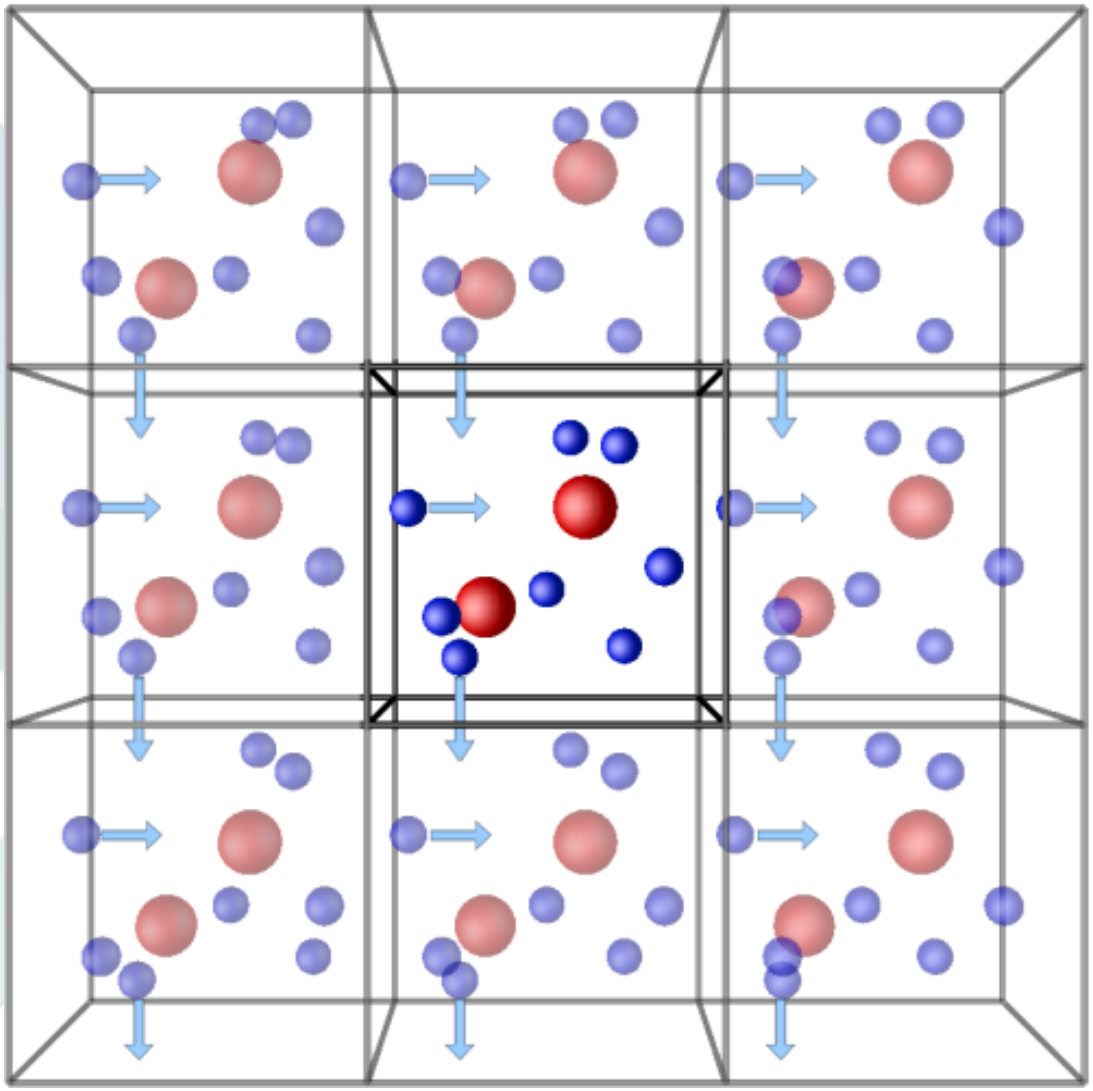
G. J. Martyna, D. J. Tobias, M. L. Klein, J. Chem. Phys 101, 1994
S. E. Feller, Y. Zhang, R. W. Pastor, B. R. Brooks, J. Chem. Phys. 103, 1995.
http://www.ks.uiuc.edu/Research/namd/2.7/ug/node32.html

Periodic Boundary Conditions

The system cannot be extended to infinity .....

What do we do if the atom leaves the simulation system?

Use Periodic Boundary Conditions (PBC).



Periodic boundary conditions

The particles are enclosed in a cell that is replicated to infinity by periodic translations. **A particle that leaves the cell on the one side is replaced by a copy entering the cell on the opposite side**, each particle is subject to the potential from all other particles in the system including images in the surrounding cells.

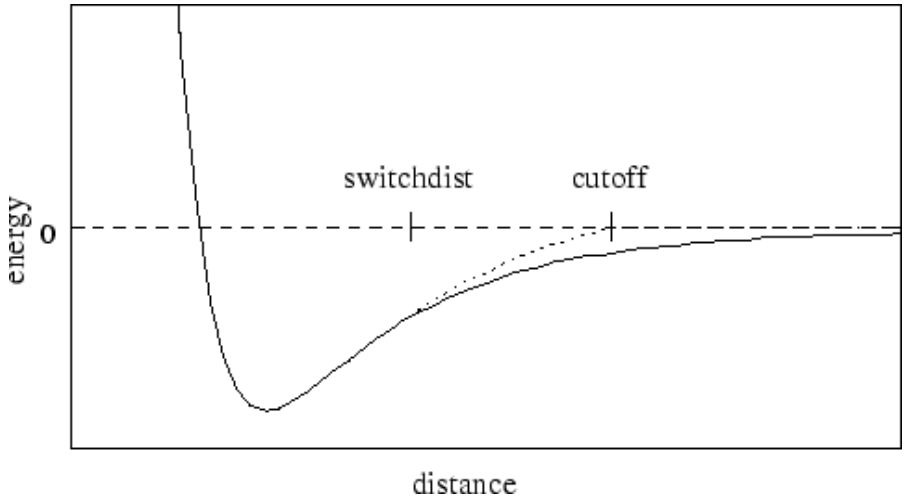
Because each cell is an identical copy of all the others all the image particles move together and need only be represented once inside the MD code.

Phillips et al, J. Comput. Chem, 26, 2005

Periodic boundary conditions

Because the vdW and electrostatic interactions exists between every nonbonded pair of atoms in the system (including those in the neighboring cells) computing the long-range interactions exactly is unfeasible. Therefore **nonbonding interactions are spatially truncated** using `cutoff` distance. The cell has to be larger then `2xcutoff`.

Electrostatic interactions can be computed fully with minimal additional cost using **Particle Mesh Ewald (PME)** method.



Phillips et al, J. Comput. Chem, 26, 2005

http://www.ks.uiuc.edu/Research/namd/2.7/ug/

Particle Mesh Ewald

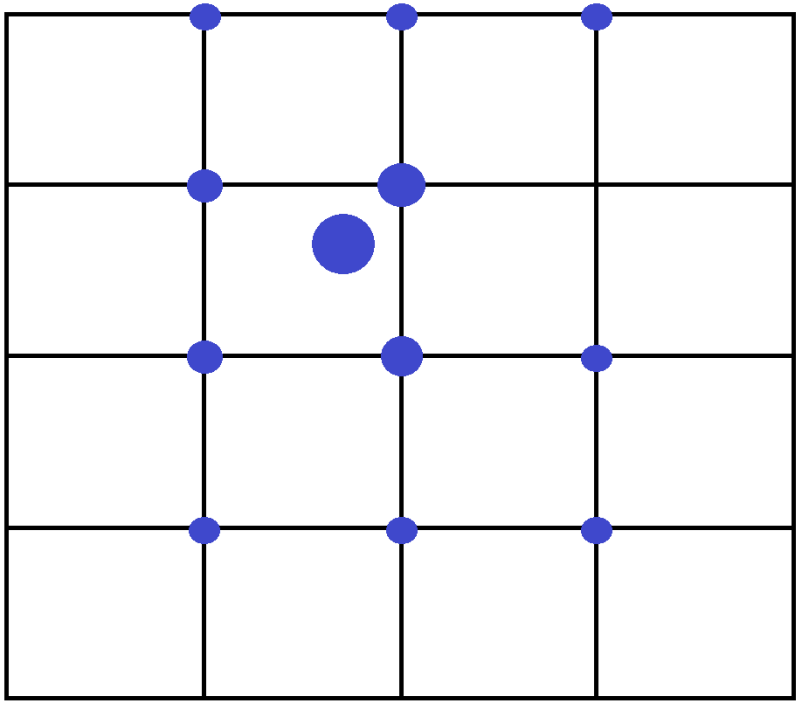
Ewald Summation is a description (introduced in 1970) of the long-range electrostatic interactions for a spatially limited system with PBC.

The direct summation of interaction energy between point particles is replaced by **two summations: a direct sum of the short-ranged potential in real space and a summation in Fourier space of the long-ranged part.** Both summations converge quickly, so they may be truncated with little loss of accuracy and great improvement in required computational time. The method uses FFT which requires that the density field be evaluated on a discrete lattice in space (mesh).

Particle Mesh Ewald

PME uses an interpolation scheme to **distribute the charges** sitting anywhere in real space **to the nodes in the uniform grid (mesh).** The particle mesh is a 3D grid created in the system over which the system charge is distributed. The weighting function is also used.

Positioning all charges on a grid enables the application of FFT and significantly reduces the computational cost.



Phillips et al, J. Comput. Chem, 26, 2005

Particle Mesh Ewald

PME method can be applied to systems with **periodic symmetry**. In MD a charge-neutral unit cell is created which is infinitely "tiled" to form images. It means that PBC are used. The unit cell must be big enough to avoid improper motion correlations between two faces of the cell and still small enough to be computationally feasible.

PME is an efficient **full electrostatic method** for use with PBC. It should not affect energy conservation, although it may affect the accuracy of the result and momentum conservation. **PME is more accurate and less expensive than larger cutoffs.** NAMD uses a PME variant called SPME.

Phillips et al, J. Comput. Chem, 26, 2005

Constraint MD

In many cases the energy transport between high-frequency (bonds) and low-frequency (twist, rotation etc) degrees of freedom is very slow. In such case finding the equilibrium state would be very time consuming. To save time **the high-frequency motions (bonds in the protein and bonds & angles in water) are frozen using geometric constrains.** It does not significantly affect the accuracy.

SHAKE algorithm satisfies bond geometry constraint in MD, and is limited to mechanical systems with a tree structure (no closed loops of constraints). Other versions: **Q-SHAKE, MSHAKE**, and **RATTLE** work well with Velocity Verlet algorithm.

RATTLE Algorithm

Constrains: the forces acting on each particle consist of physical force and forces resulting from constrains.

Holonomic constrains σ^{ij} between particles i and j depend on their positions x_i and x_j and the desired bond length d_{ij}, but don't depend on time):

$$\sigma^{ij}(x) = \left\| x_i - x_j \right\|^2 - d_{ij}^2 = 0$$

Again, we have to modify the integration algorithm -> RATTLE algorithm.

RATTLE Algorithm

Understanding of the algorithm would require using advanced equations (Hamilton equation) so the RATTLE algorithm is not explicitly presented.

RATTLE: The next position of atom has to be a solution of Newton equation and simultaneously satisfy the constraining conditions.

We need to calculate the constraining force and check who is constrained and who is not -> more to calculate. But later on we don't have to include all terms and atoms in the next step of the Newton equation integration -> **calculation speed-up** ☺

Steered Molecular Dynamics

Biologically important events often involve transitions from one equilibrium state to another, such as the binding or dissociation of a ligand. Such processes involve **the rare event of barrier crossing and are difficult to reproduce on MD timescales**.

The idea behind **SMD is to apply the external force to guide the system from one state to another.**

Experimental techniques as **AFM (atomic force microscopy) and optical tweezers** have enhanced understanding of the mechanical properties of biopolymers. **SMD is the *in silico* complement of those techniques.**

Phillips et al, J. Comput. Chem, 26, 2005

SMD

Within SMD **external force(s) are applied to molecules in simulation to probe their mechanical properties and accelerate processes that are otherwise too slow to model**.

The external force can be applied to one or more atoms which are called SMD atoms. Simultaneously, another atoms can be kept fixed.

There are two types of SMD: **constant force pulling and constant velocity pulling.**

Phillips et al, J. Comput. Chem, 26, 2005

SMD

In constant-force SMD the atoms to which the force is applied are a subject to a fixed, constant force in addition to the Force Field potential. Typical applied forces range from tens to a thousand pN.

Constant velocity SMD simulates the action of a moving AFM cantilever on a protein. **An atom** of the protein or a center of mass of a group of atoms **is harmonically restrained to a point in space that is then shifted in a chosen direction** at a predetermined constant velocity, forcing the restrained atoms to follow. This variant seems to be **more interesting** and easier compared with the experimental results.

Phillips et al, J. Comput. Chem, 26, 2005

SMD

In constant velocity pulling **the SMD atom (pulled atom) is attached to a dummy atom via a virtual spring**. This dummy atom is moved with a constant velocity and the force between them is measured:

$$\vec{F}_{SMD}(\vec{r}) = -\nabla U_{SMD}(\vec{r})$$

$$U_{SMD}(\vec{r}) = \frac{1}{2} k\left[vt - (\vec{r} - \vec{r}_0)\vec{n}\right]^2$$

Where: k - spring constant

v - pulling velocity

$\vec{r}$ - actual position of SMD atom

$\vec{r}_0$ - initial position of SMD atom

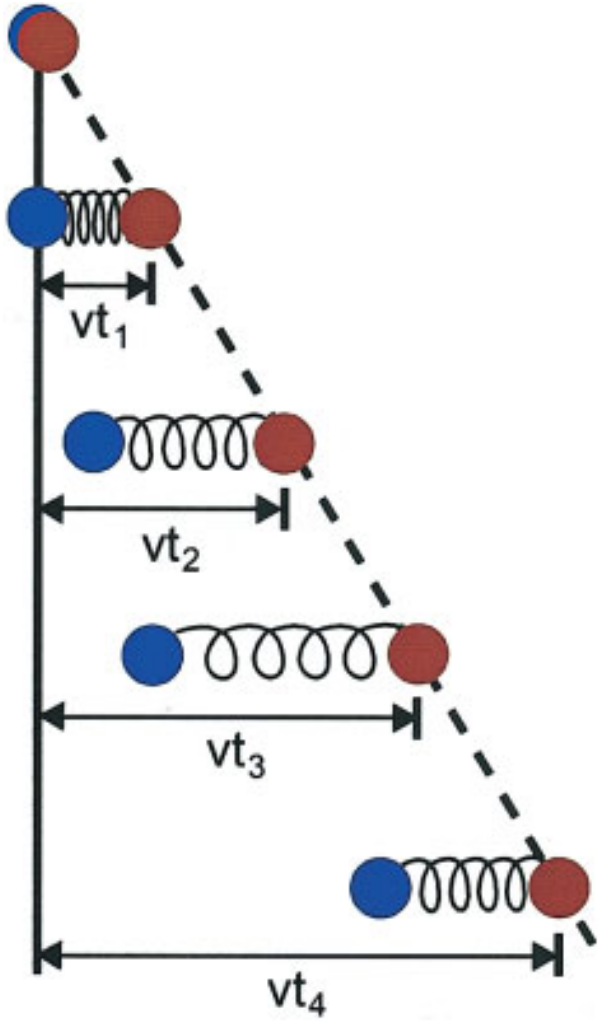
$\vec{n}$ - pulling direction

Phillips et al, J. Comput. Chem, 26, 2005

The U_{SMD} has to be added to the force field potential.

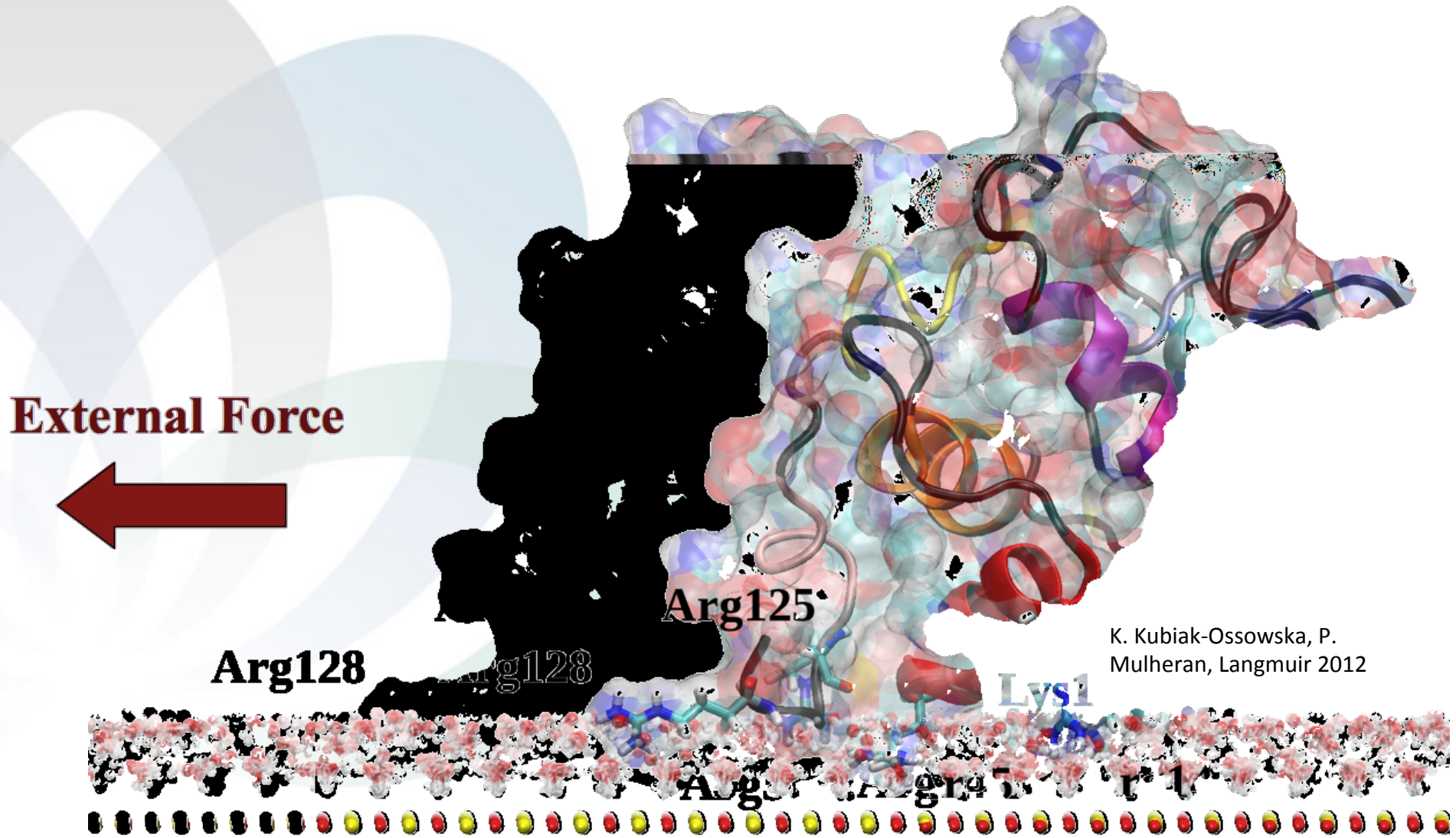
SMD

Constant velocity SMD. The dummy atom is shown in red and the SMD atom in blue. **As the dummy atom moves at constant velocity the SMD atom experiences a force** that depends linearly on the distance between both atoms.



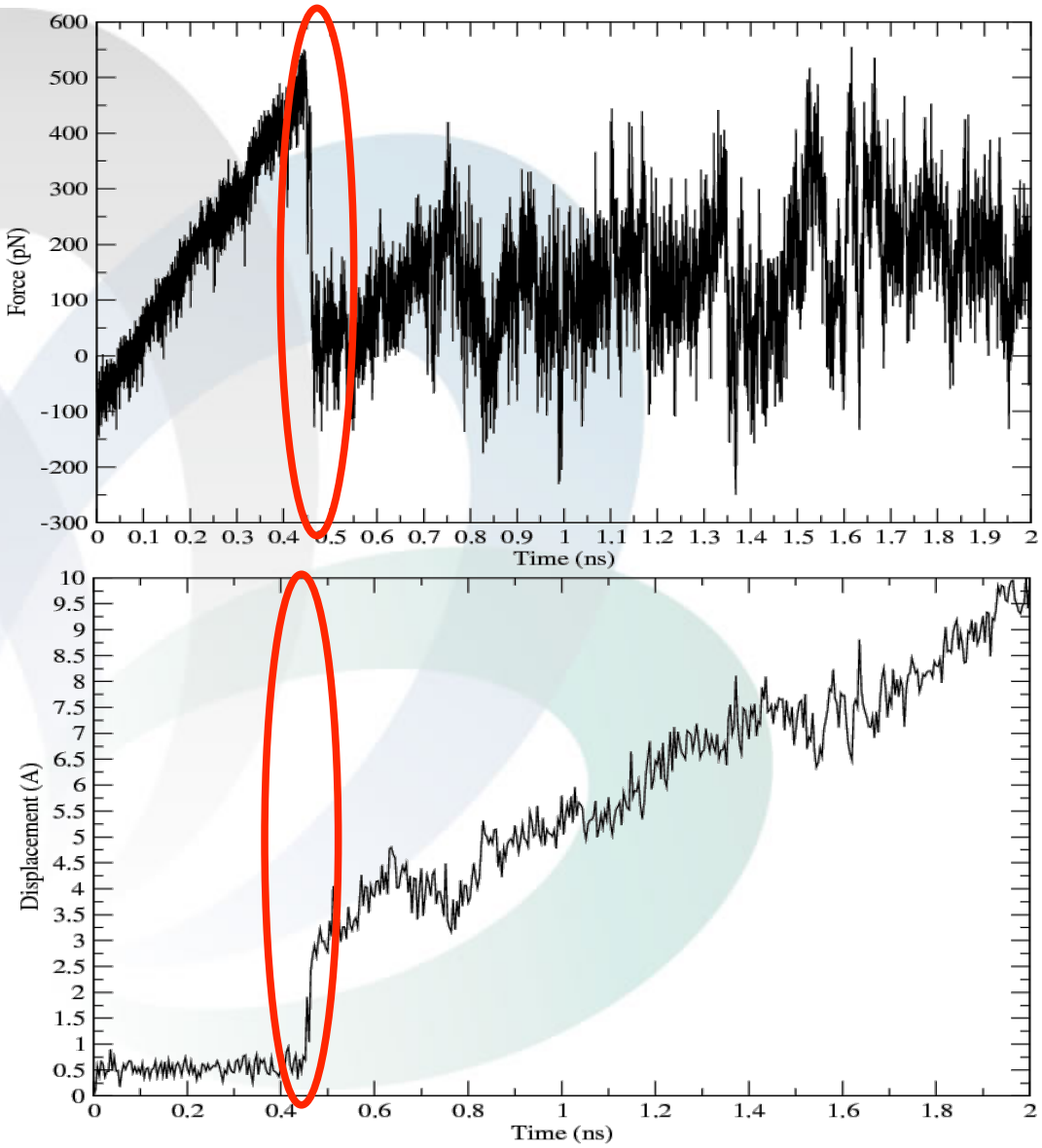
Phillips et al, J. Comput. Chem, 26, 2005

SMD



K. Kubiak-Ossowska, P. Mulheran, Langmuir 2012

SMD



$$dE = \left(F_0 + \frac{dF}{2}\right)\frac{dF}{k}$$

Where:

F_0 - the force at the end of transition,

k – spring constant

dF – the force change

Here:

F_0 ~ 0 N

dF ~ 575 pN

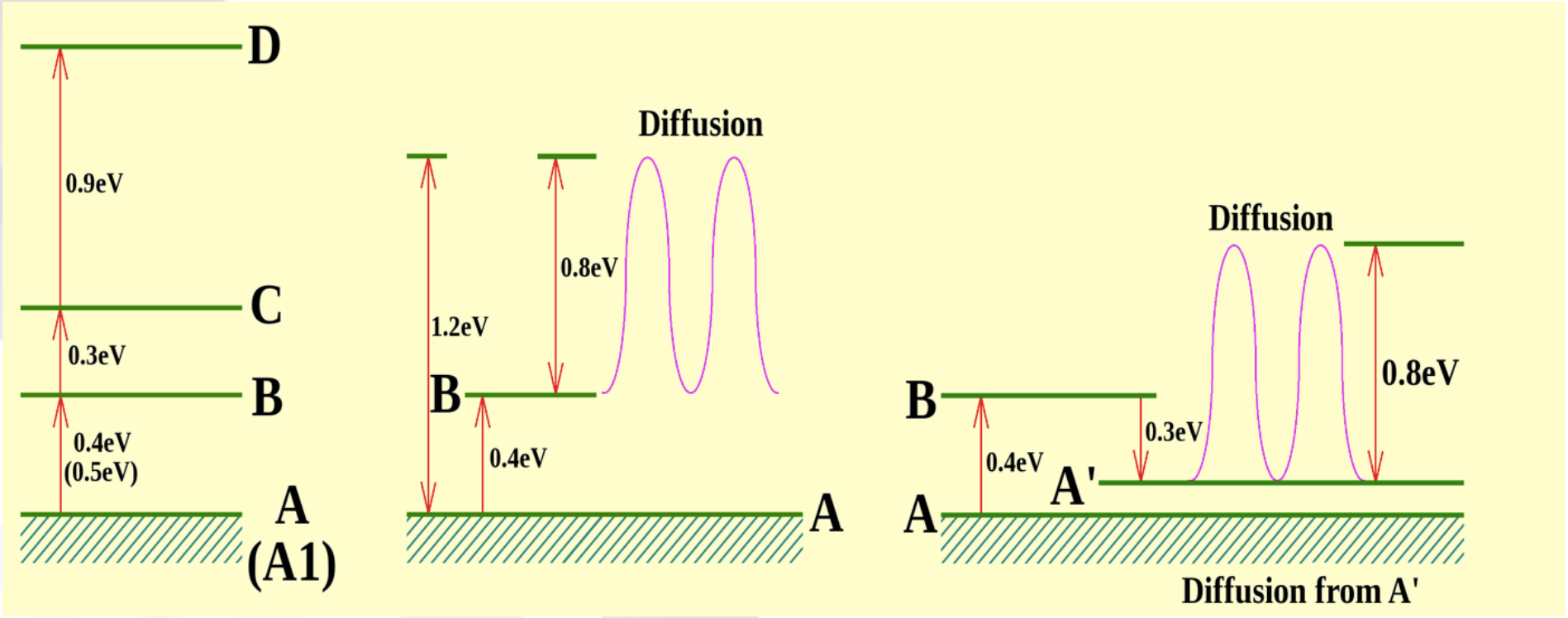
k = 278 Å/pN

dE=543pN*Å (10^{-22}J)

dE=0.34 eV

K. Kubiak-Ossowska, P. Mulheran, Langmuir 2012

Steered Molecular Dynamics



K. Kubiak-Ossowska, P. Mulheran, Langmuir 2012

MD limitations

We would like to simulate bigger and **bigger systems for longer timescales** relevant to the timescales of natural processes. But ….

- The bigger the system – the longer computation time
- The longer simulation timescale – the longer computation time.
- Explicit water – the longer computation time

MD simulations are computationally costly, they require hundreds or even millions of CPU years…..

Typical length of an MD run is now in the order of hundreds of nanoseconds, sometimes reaching microseconds.

Cost reduction

Methods to reduce the computational cost:

- **Cutoff:** The most time-consuming part in MD are calculations of the $U_{nonbonded}$ interactions (vdW and electrostatics interactions) which should be calculated to the infinitum… **The cutoff distance used is 12 Å**.

- **Constrained MD**
  a) By omitting high frequency terms (O-H bond stretching) using **RATTLE** the time step can be increased to 2fs without losing accuracy.
  b) Water molecules are kept rigid.

Cost reduction

- **Implicit solvent:** popular years ago, now not really due to known role of the interactions with water, H-bonds etc
- **SMD**
- **Correct preparation** of the system: for example omitting the diffusion and guessing of the initial orientation
- **H**igh **P**erformance **C**omputers (**HPC**s)

Further reading

1. J. C. Phillips et all, Scalable Molecular Dynamics with NAMD, J. Comput. Chem. 2005
2. M. P. Allen, D. J. Tildesley, Computer Simulation of Liquids, Oxford University Press, 2007
3. M. Griebel, G. Zumbusch, S. Knapek, Numerical Simulation in Molecular Dynamics
4. O. M Becker, M. Karplus, A guide to Biomolecular Simulations, 2006
5. http://www.ks.uiuc.edu/Research/namd/
6. http://www.ks.uiuc.edu/Research/namd/2.7/ug/ug.html

Reusing this material

archer

epcc

THE UNIVERSITY OF EDINBURGH