

# **Molecular Dynamics**

# Molecular Dynamics

- is computer simulation of physical movements by atoms and molecules
- is frequently used in the study of proteins and biomolecules, as well as in materials science
- is a specialized discipline of molecular modeling and computer simulation based on statistical mechanics

Wikipedia

# Molecular Dynamics

Biological molecules exhibit a wide range of time scales over which specific processes occur; for example

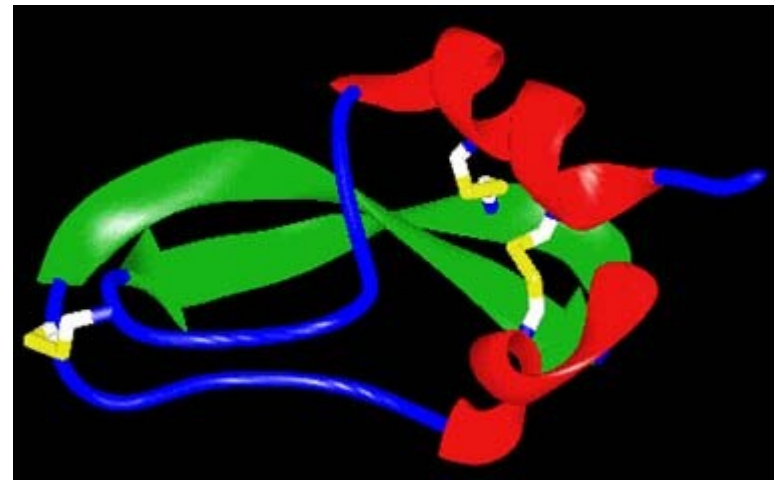
- **Local Motions (0.01 to 5 Å,  $10^{-15}$  to  $10^{-1}$  s)**
  - Atomic fluctuations
  - Sidechain Motions
  - Loop Motions
- **Rigid Body Motions (1 to 10Å,  $10^{-9}$  to 1s)**
  - Helix Motions
  - Domain Motions (hinge bending)
  - Subunit motions
- **Large-Scale Motions ( $> 5\text{Å}$ ,  $10^{-7}$  to  $10^4$  s)**
  - Helix coil transitions
  - Dissociation/Association
  - Folding and Unfolding

# Molecular Dynamics

- Molecular dynamics simulations permit the study of complex, dynamic processes that occur in biological systems. These include, for example,
  - Protein stability
  - Conformational changes
  - Protein folding
  - Molecular recognition: proteins, DNA, membranes, complexes
  - Ion transport in biological systems
- and provide the mean to carry out the following studies,
  - Drug Design
  - Structure determination: X-ray and NMR

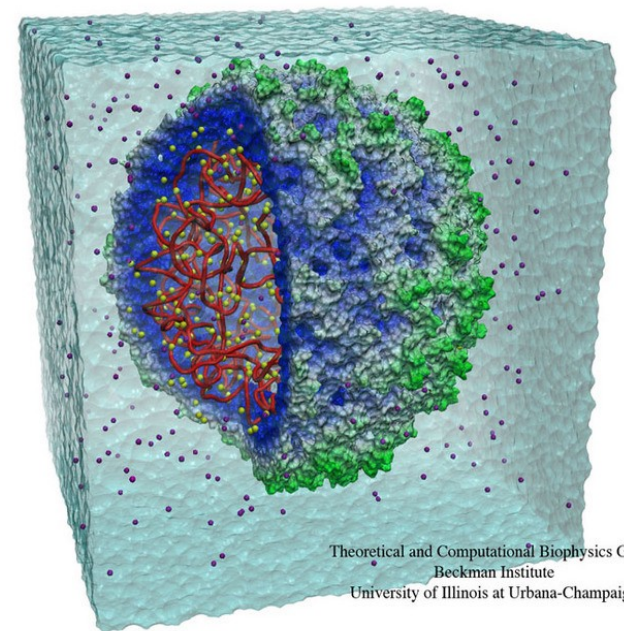
# Molecular Dynamics

- First macromolecular MD simulation published (1977, Size: 500 atoms, Simulation Time: 9.2 ps=0.0092 ns, Program: CHARMM precursor) Protein: Bovine Pancreatic Trypsine Inhibitor. This is one of the best studied proteins in terms of folding and kinetics. Its simulation published in Nature magazine paved the way for understanding protein motion as essential in function and not just accessory.



# Molecular Dynamics

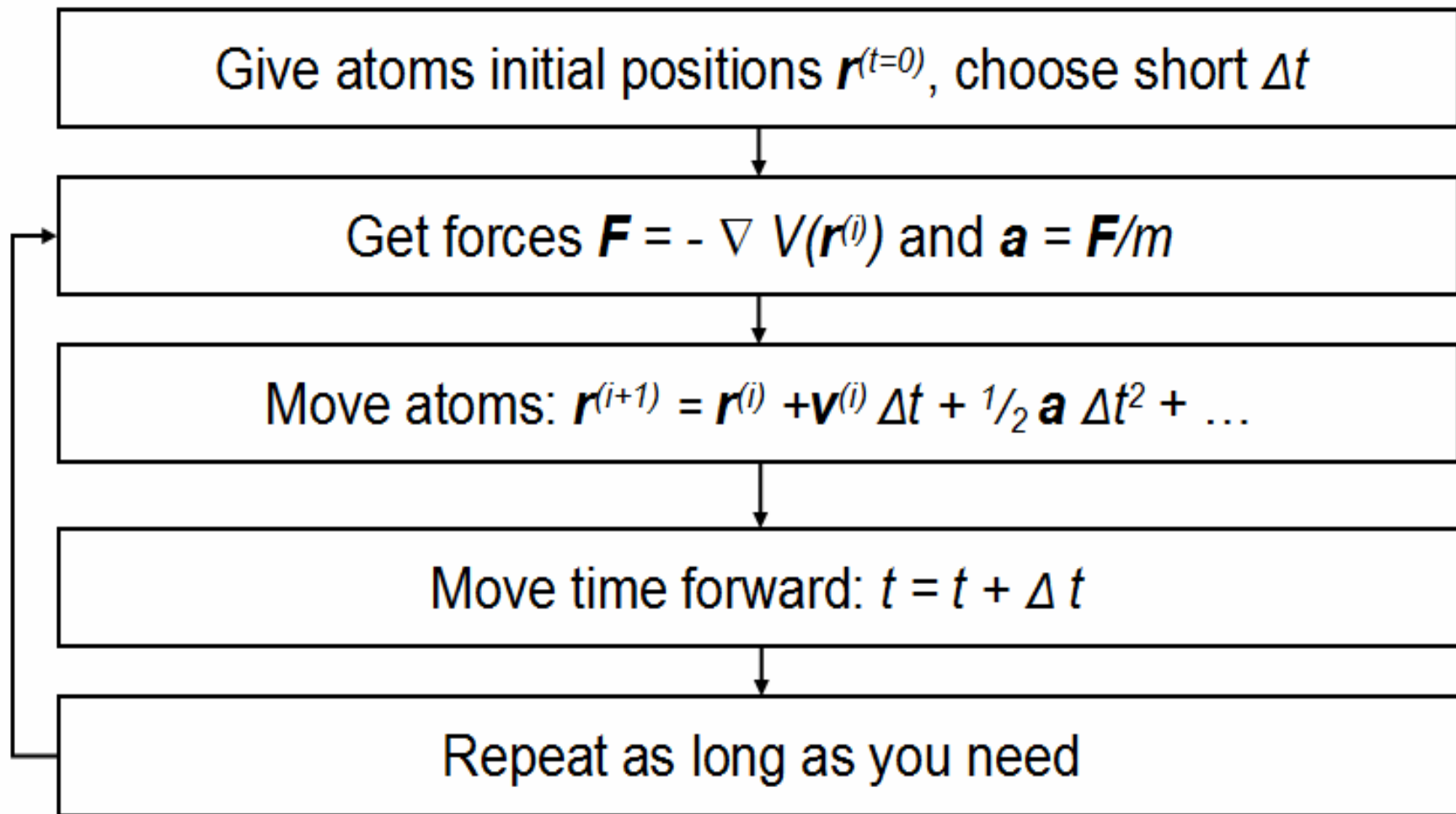
- MD simulation of the complete satellite tobacco mosaic virus (STMV) (2006, Size: 1 million atoms, Simulation time: 50 ns, program: NAMD) This virus is a small, icosahedral plant virus which worsens the symptoms of infection by Tobacco Mosaic Virus (TMV). Molecular dynamics simulations were used to probe the mechanisms of viral assembly. The entire STMV particle consists of 60 identical copies of a single protein that make up the viral capsid (coating), and a 1063 nucleotide single stranded RNA genome. One key finding is that the capsid is very unstable when there is no RNA inside. The simulation would take a single 2006 desktop computer around 35 years to complete. It was thus done in many processors in parallel with continuous communication between them



# MD ensembles

- Microcanonical ensemble (**NVE**) : The thermodynamic state characterized by a fixed number of atoms,  $N$ , a fixed volume,  $V$ , and a fixed energy,  $E$ . This corresponds to an isolated system.
- Canonical Ensemble (**NVT**): This is a collection of all systems whose thermodynamic state is characterized by a fixed number of atoms,  $N$ , a fixed volume,  $V$ , and a fixed temperature,  $T$ .
- Isobaric-Isothermal Ensemble (**NPT**): This ensemble is characterized by a fixed number of atoms,  $N$ , a fixed pressure,  $P$ , and a fixed temperature,  $T$ .
- Grand canonical Ensemble ( **$\mu$ VT**): The thermodynamic state for this ensemble is characterized by a fixed chemical potential,  $\mu$ , a fixed volume,  $V$ , and a fixed temperature,  $T$ .

# MD – how it works

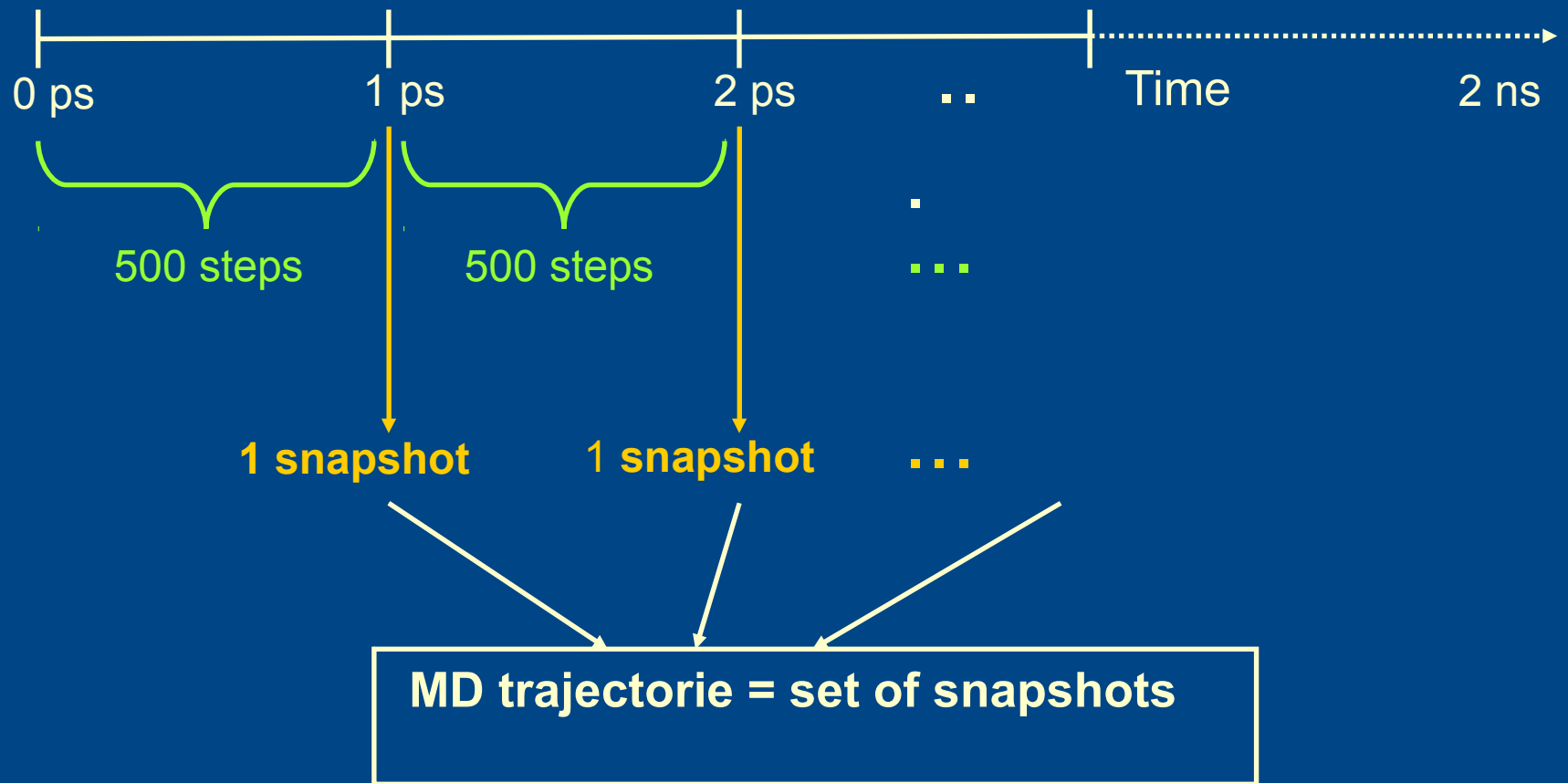




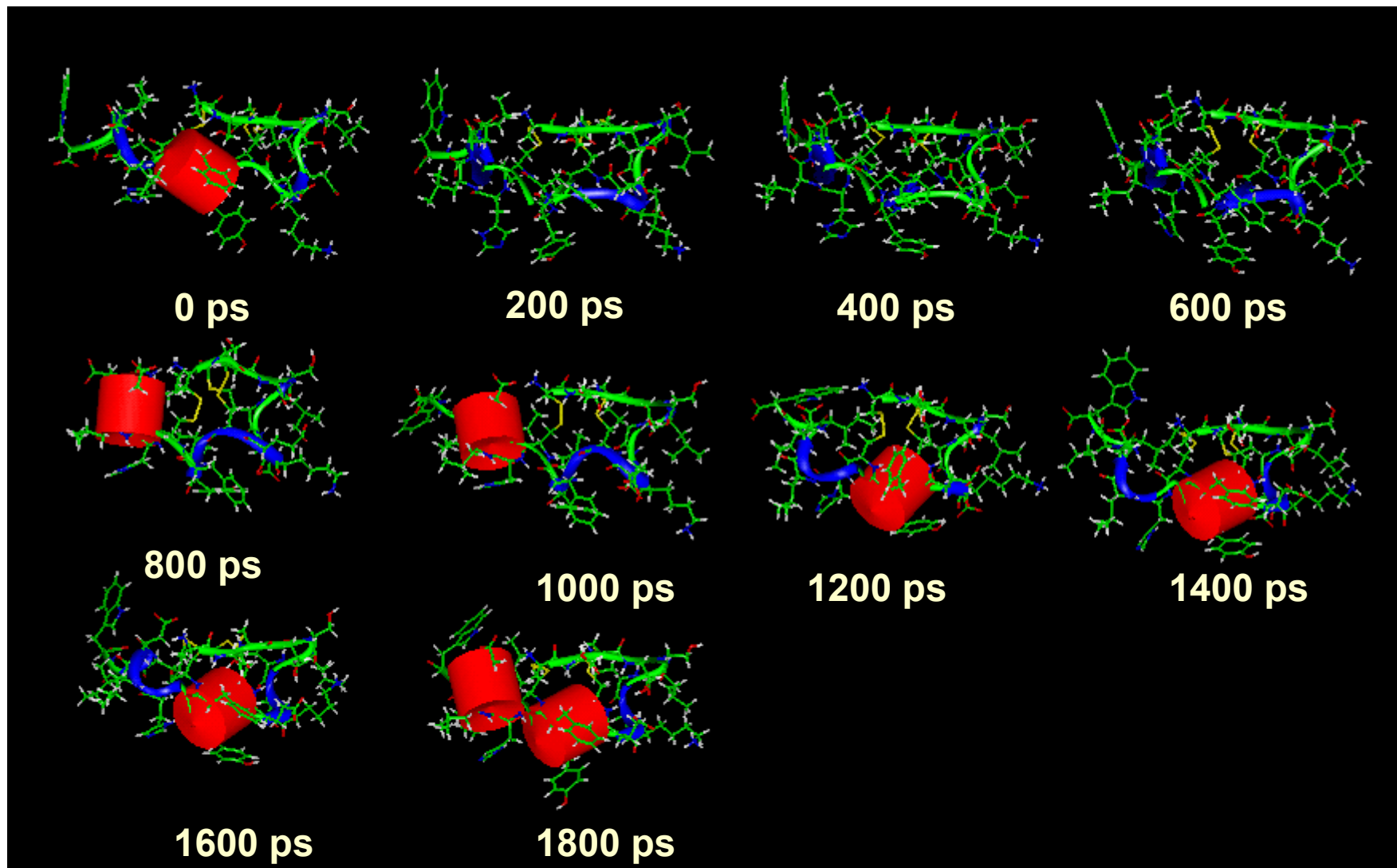
# Molecular Dynamics

1 MD step = 0.002 ps = 2 fs

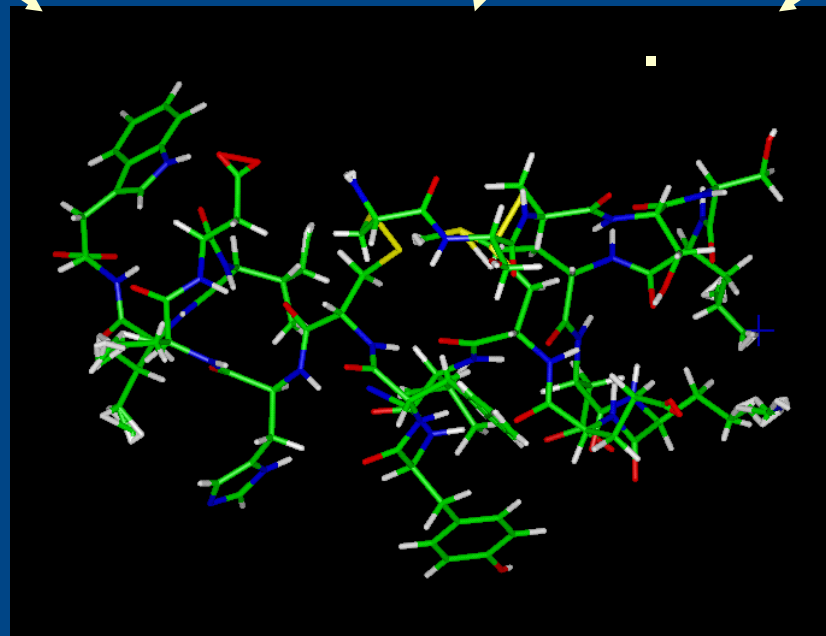
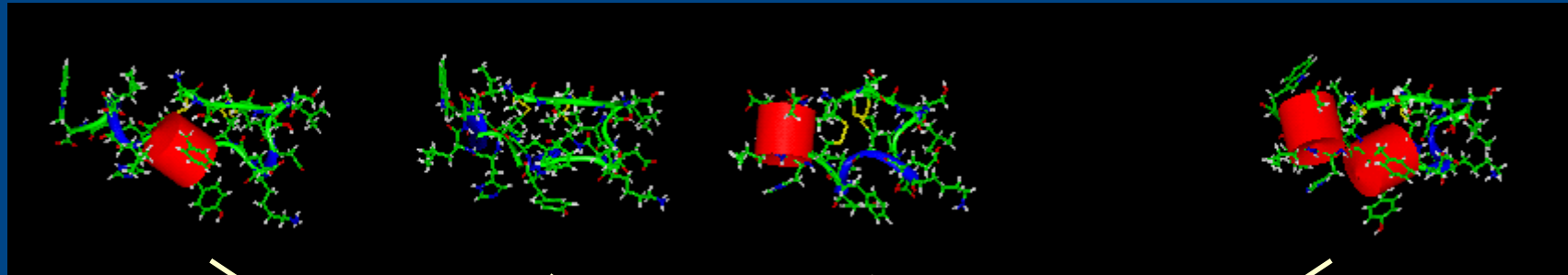
Saved each 500 step



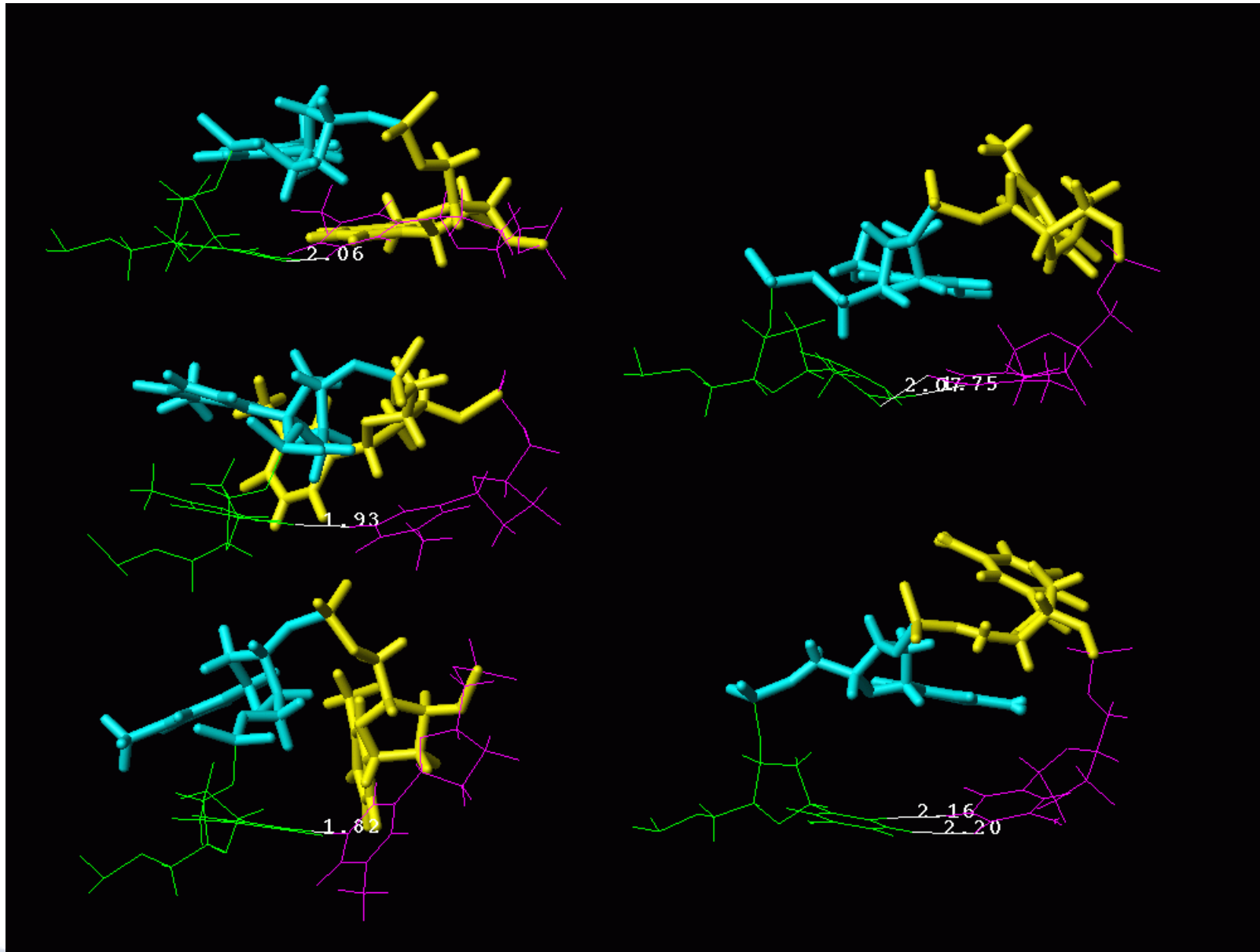
# Molecular Dynamics - trajectory



# MD – average structure



# MD – structural changes



# MD – math behind

Newton's equation of motion is given by

$$F_i = m_i a_i$$

where  $F_i$  is the force exerted on particle  $i$ ,  $m_i$  is the mass of particle  $i$  and  $a_i$  is the acceleration of particle  $i$ . The force can also be expressed as the gradient of the potential energy,

$$F_i = -\nabla_i V$$

Combining these two equations yields

$$-\frac{dV}{dr_i} = m_i \frac{d^2 r_i}{dt^2}$$

where  $V$  is the potential energy of the system. Newton's equation of motion can then relate the derivative of the potential energy to the changes in position as a function of time.

# MD – math behind

The potential energy is a function of the atomic positions ( $3N$ ) of all the atoms in the system. Due to the complicated nature of this function, there is **no analytical solution** to the equations of motion; they must be solved numerically.

Numerous numerical algorithms have been developed for integrating the equations of motion.

- \* **Verlet algorithm**
- \* **Leap-frog algorithm**
- \* **Velocity Verlet**
- \* **Beeman's algorithm**

# MD – Verlet algorithm

All the integration algorithms assume the positions, velocities and accelerations can be approximated by a Taylor series expansion:

$$r(t + \delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 + \dots$$

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

$$a(t + \delta t) = a(t) + b(t)\delta t + \dots$$

# MD – Verlet algorithm

To derive the Verlet algorithm one can write

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

$$r(t - \delta t) = r(t) - v(t)\delta t + \frac{1}{2}a(t)\delta t^2$$

Summing these two equations, one obtains

$$r(t + \delta t) = 2r(t) - r(t - \delta t) + a(t)\delta t^2$$

The Verlet algorithm uses positions and accelerations at time  $t$  and the positions from time  $t-\delta t$  to calculate new positions at time  $t+\delta t$ . The Verlet algorithm uses no explicit velocities. The advantages of the Verlet algorithm are, i) it is straightforward, and ii) the storage requirements are modest. The disadvantage is that the algorithm is of moderate precision.



# MD - The Velocity Verlet algorithm

This algorithm yields positions, velocities and accelerations at time  $t$ . There is no compromise on precision.

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

$$v(t + \delta t) = v(t) + \frac{1}{2}[a(t) + a(t + \delta t)]\delta t$$

# MD – Beeman's algorithm

This algorithm is closely related to the Verlet algorithm

$$r(t + \delta t) = r(t) + v(t)\delta t + \frac{2}{3}a(t)\delta t^2 - \frac{1}{6}a(t - \delta t)\delta t^2$$

$$v(t + \delta t) = v(t) + v(t)\delta t + \frac{1}{3}a(t)\delta t + \frac{5}{6}a(t)\delta t - \frac{1}{6}a(t - \delta t)\delta t$$

The advantage of this algorithm is that it provides a more accurate expression for the velocities and better energy conservation. The disadvantage is that the more complex expressions make the calculation more expensive.

# MD – The leap-frog algorithm

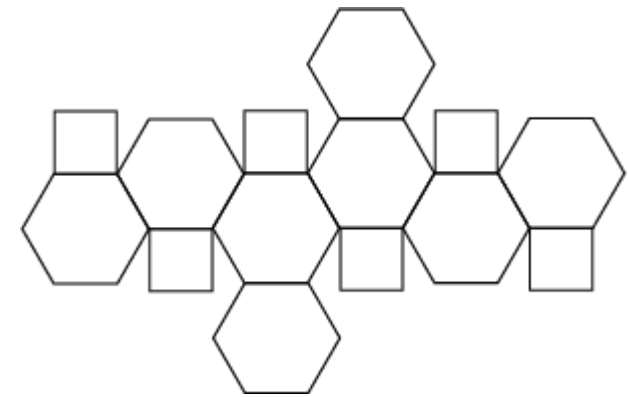
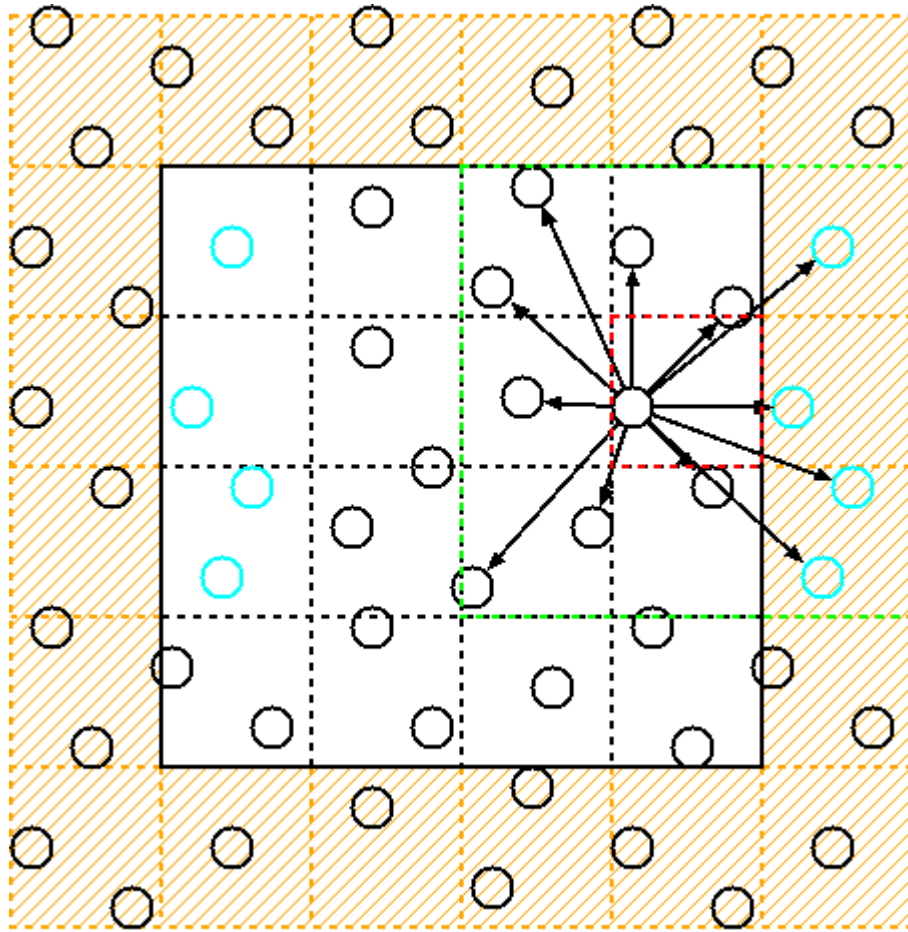
$$r(t + \delta t) = r(t) + v\left(t + \frac{1}{2}\delta t\right)\delta t$$

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

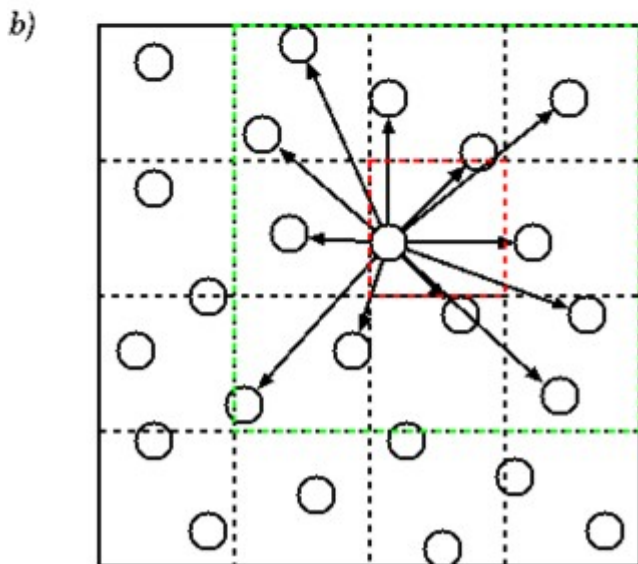
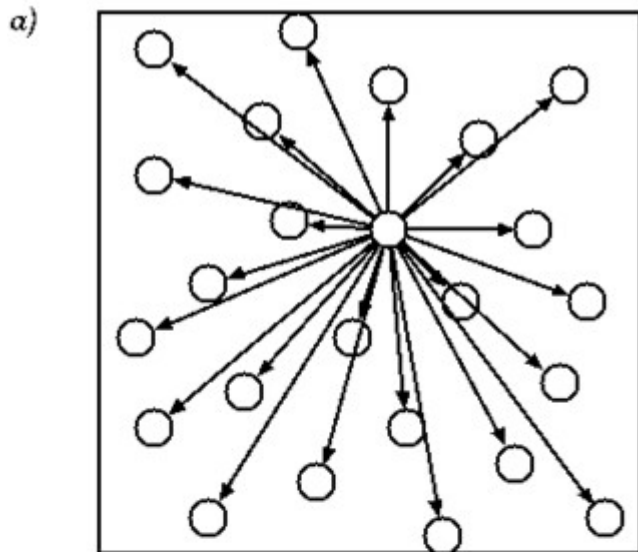
In this algorithm, the velocities are first calculated at time  $t + 1/2\delta t$ ; these are used to calculate the positions,  $r$ , at time  $t + \delta t$ . In this way, the velocities leap over the positions, then the positions leap over the velocities. The advantage of this algorithm is that the velocities are explicitly calculated, however, the disadvantage is that they are not calculated at the same time as the positions. The velocities at time  $t$  can be approximated by the relationship:

$$v(t) = \frac{1}{2} \left[ v\left(t - \frac{1}{2}\delta t\right) + v\left(t + \frac{1}{2}\delta t\right) \right]$$

# MD – Periodic boundary conditions



# MD – short range interactions



The pairwise interactions for a single particle can be computed by a) computing the interaction to all other particles or b) by dividing the domain into cells with an edge length of at least the cut-off radius of the interaction potential and computing the interaction between the particle and all particles in the same (red) and in the adjacent (green) cells.

# MD – Long range interactions

- Particle Mesh Evald

$$E_{\text{TOT}} = \sum_{i,j} \varphi(\mathbf{r}_j - \mathbf{r}_i) = E_{sr} + E_{\ell r}$$

$$E_{sr} = \sum_{i,j} \varphi_{sr}(\mathbf{r}_j - \mathbf{r}_i)$$

**Real space**

$$E_{\ell r} = \sum_{\mathbf{k}} \tilde{\Phi}_{\ell r}(\mathbf{k}) |\tilde{\rho}(\mathbf{k})|^2$$

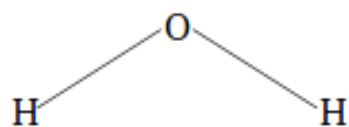
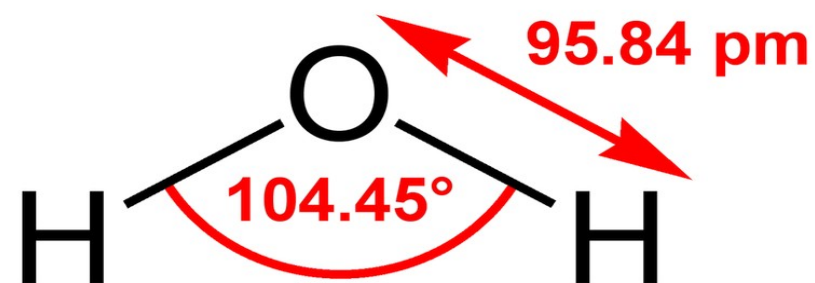
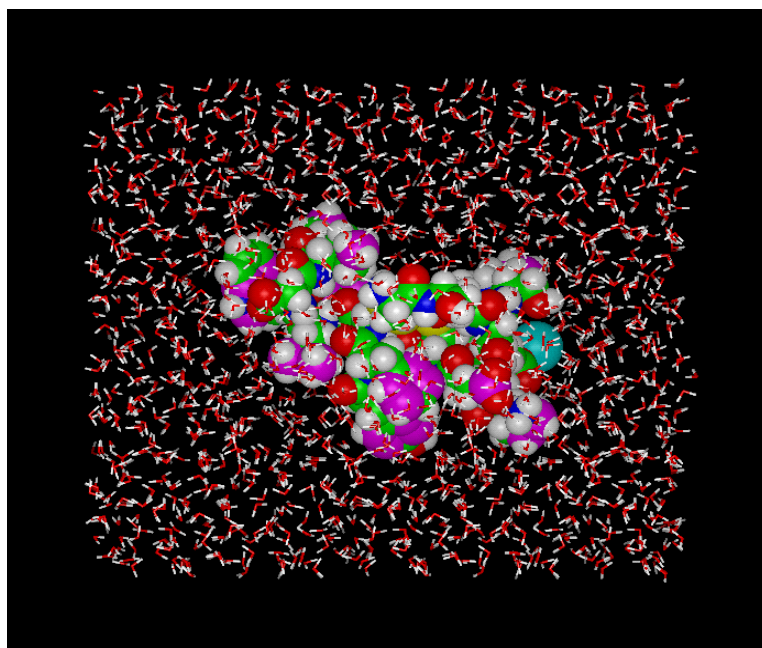
**Fourier space**

# MD – Solvation model

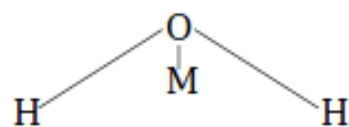
- Implicit solvent
  - a method of representing solvent as a continuous medium instead of individual “explicit” solvent molecules
  - Poisson – Boltzmann model
  - Generalized Born model

# MD – Solvation model

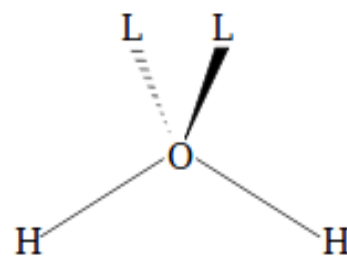
- Explicit solvent model



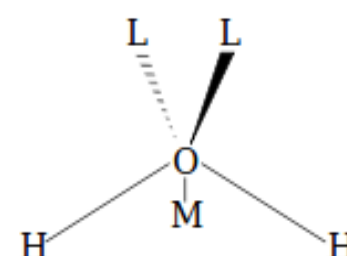
3-site



4-site



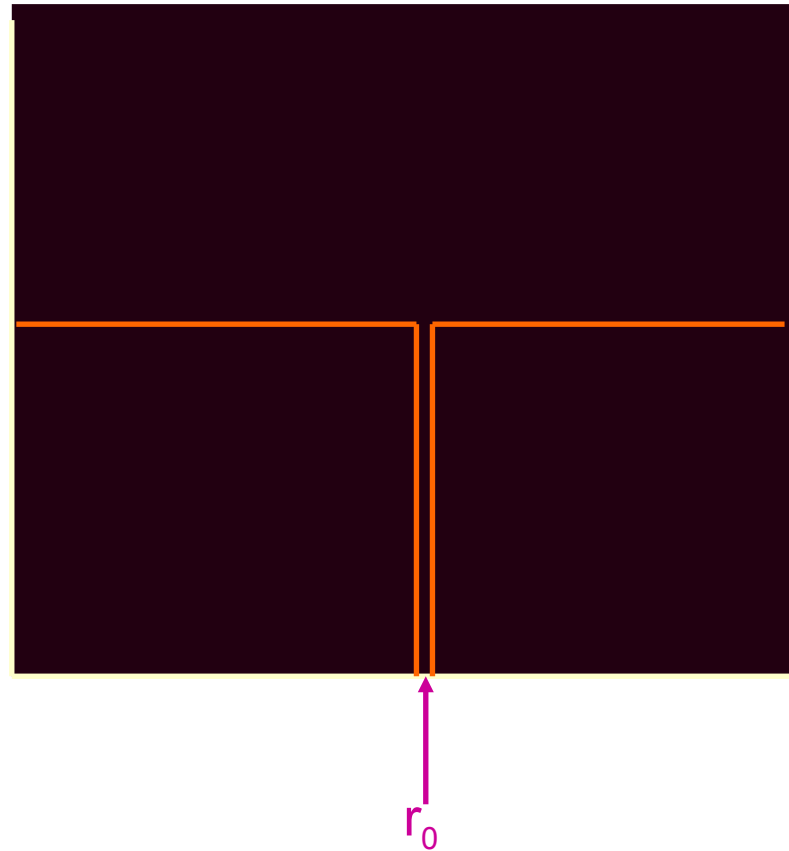
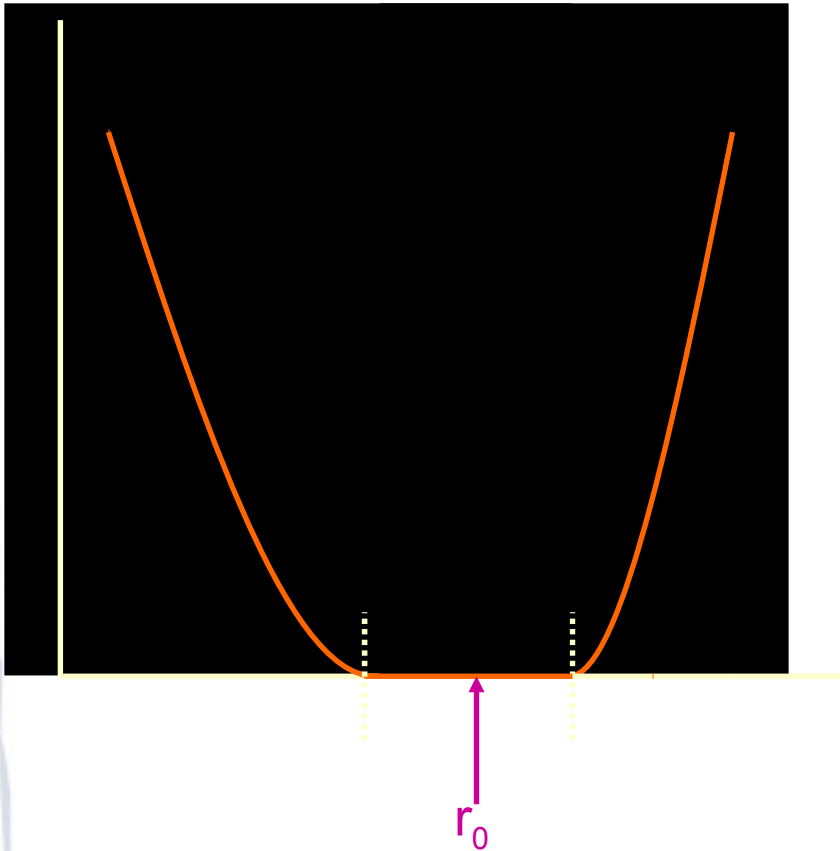
5-site



6-site

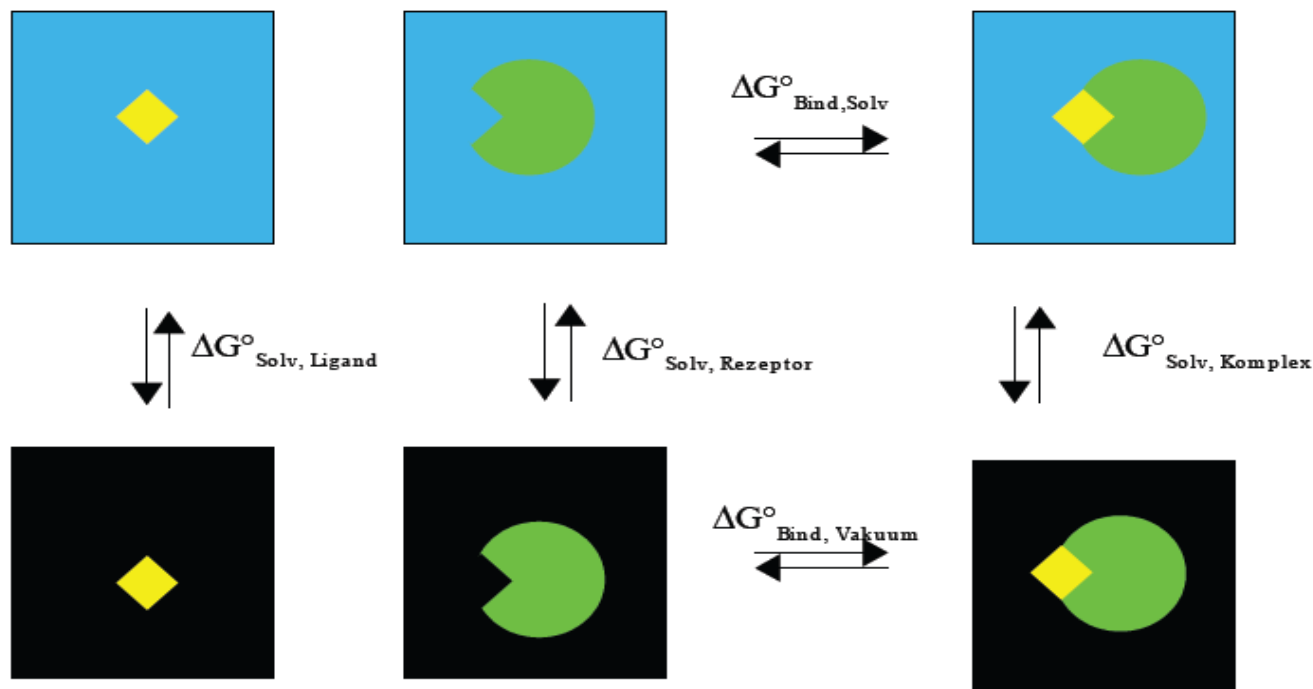


# Restrains x Constraints



# MD – post-simulation analysis

- MM-PBSA , MM-GBSA analysis



$$\Delta G^{\circ}_{\text{Bind,Solv}} = \Delta G^{\circ}_{\text{Bind,Vakuum}} + \Delta G^{\circ}_{\text{Solv,Komplex}} - \left( \Delta G^{\circ}_{\text{Solv,Ligand}} + \Delta G^{\circ}_{\text{Solv,Rezeptor}} \right)$$

# MD – post-simulation analysis

- Secondary structure of proteins – DSSP

- DSSP recognizes eight types of secondary structure (each identified by its own symbol), depending on the pattern of hydrogen bonds. The 3<sub>10</sub> helix, alpha helix and pi helix are symbolized as G, H and I, respectively, and are recognized by having a repetitive sequence of hydrogen bonds in which the donor residue is three, four, or five residues later in the backbone. DSSP recognizes two types of hydrogen-bond pairs in beta sheet structures, the parallel and antiparallel bridge; a lone bridge is symbolized by B (a beta bridge), whereas longer sets of H-bond pairs of the same type are symbolized by E for extended, which is equivalent to a beta sheet. Sheets with beta bulges are likewise symbolized by E. The remaining types are T for turn (featuring a hydrogen bond typical of a helix), S for a region of high curvature (if the angle between the vector from C $\alpha_i$  to C $\alpha_{i+2}$  and the vector from C $\alpha_{i-2}$  to C $\alpha_i$  is less than 70°, and a blank if no other rule pertains. Additionally there is L meaning "loop" or "other"

# MD - software

- AMBER
- NAMD
- GROMOS
- GROMACS
- YASARA
- CHARMM