CHEM-E4115
Computational Chemistry I (5op)
2nd part: molecular modelling

Book Chapters 4.1-4.7

Empirical Force Field Models: Molecular Mechanics

Revision: From quantum mechanics to molecular mechanics

- Quantum mechanics: Schrödinger's equation

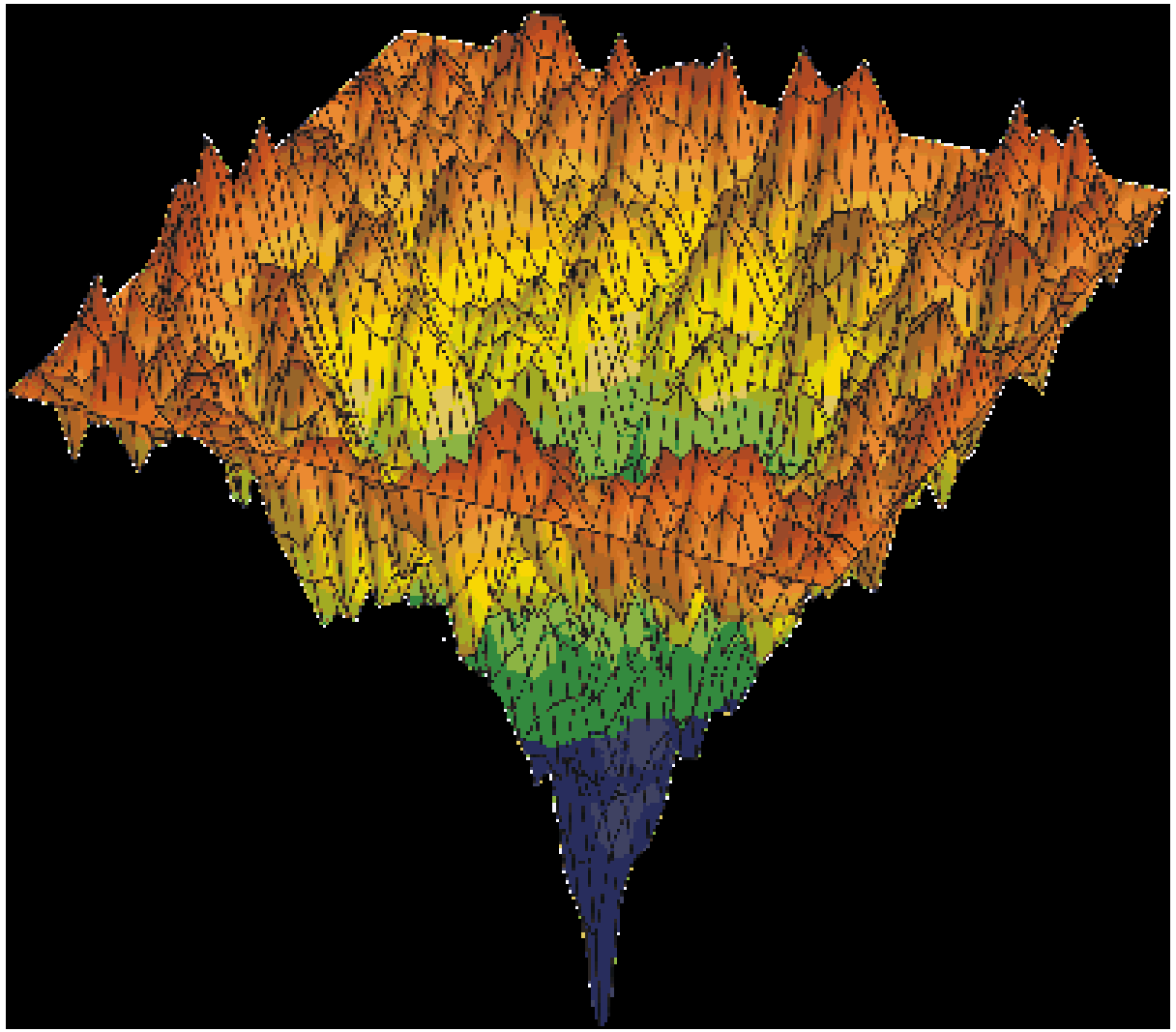
$$i\hbar \frac{\partial}{\partial t} \psi = \hat{H}\psi$$

Time-dependent Schrödinger equation (general)

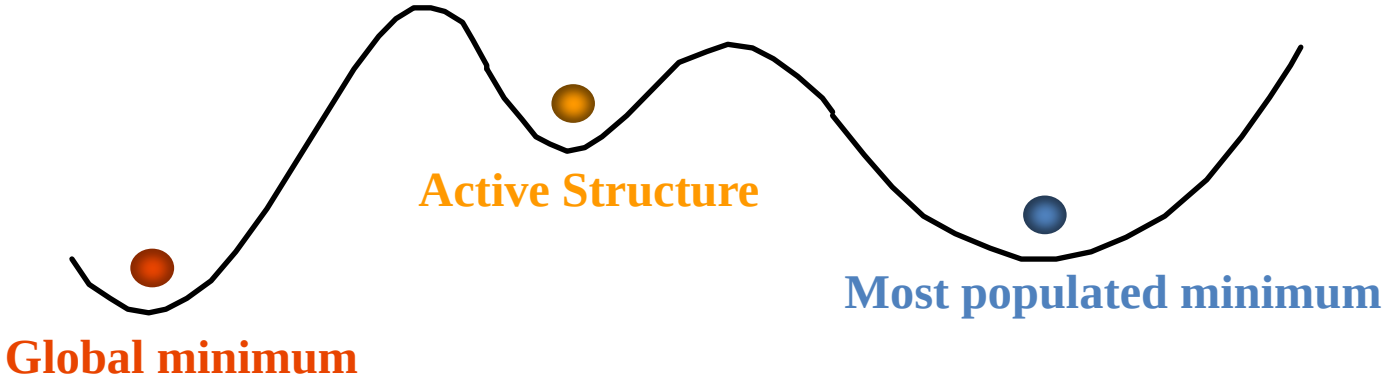
- Born-Oppenheimer approximation:
  - Electrons and nuclei motion treated separately
  - Consequently energy of a molecule in ground (electronic) state can be considered as a function of the nuclear coordinates only
  - If one or several of the nuclei move, this energy changes

→ Potential energy surface

This lecture: Potential energy surface, and basics of sampling it; measurable quantities



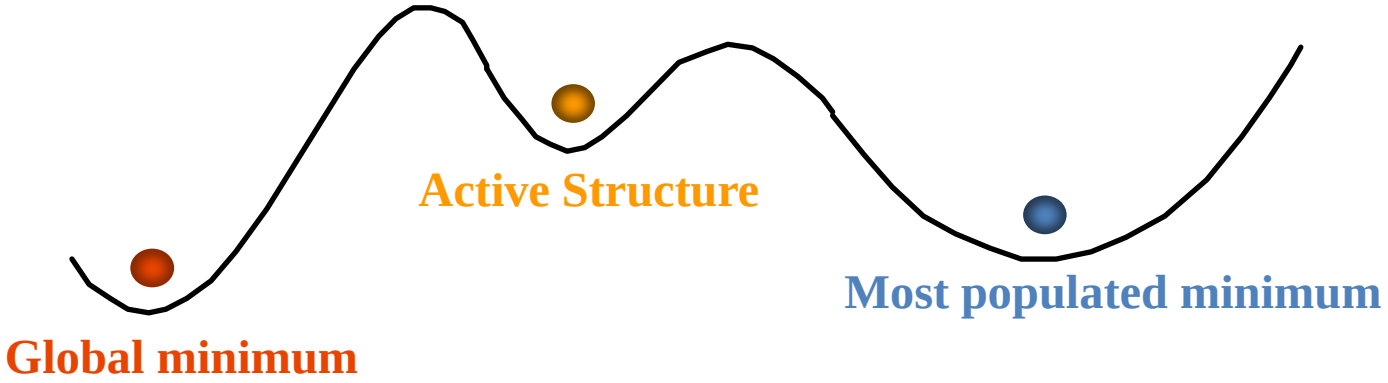
Molecular modelling and simulations



Potential energy surface

- Interested in
  - Global minimum molecular conformation / system configuration
  - Active structures
  - Relative populations
  - Transitions and transition pathways between the states

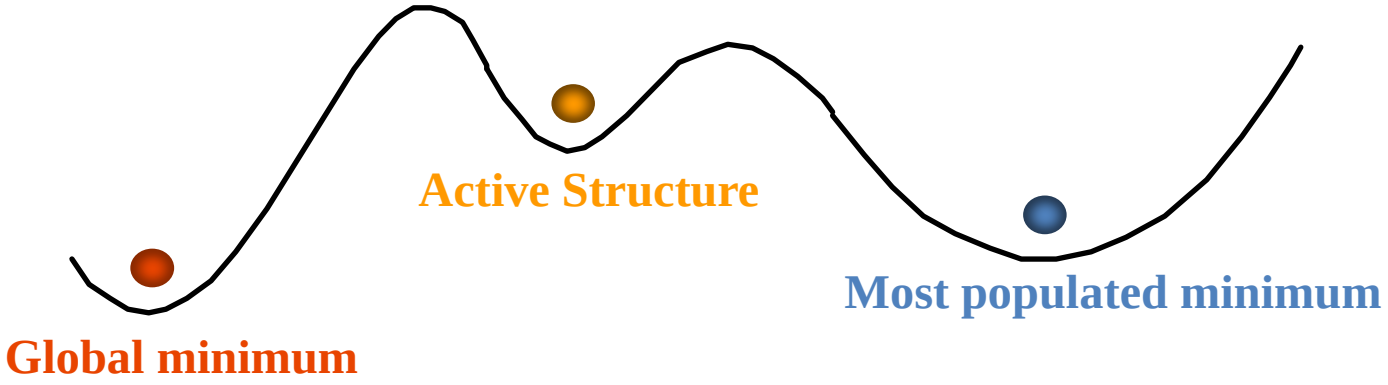
Molecular modelling and simulations: How to resolve?



Potential energy surface

- Typical methodology
  - Minimum energy configuration determination (no dynamics, just potential energy surface)
  - Molecular dynamics (deterministic dynamic sampling of the potential energy surface)
  - Monte Carlo (stochastic sampling of the potential energy surface)

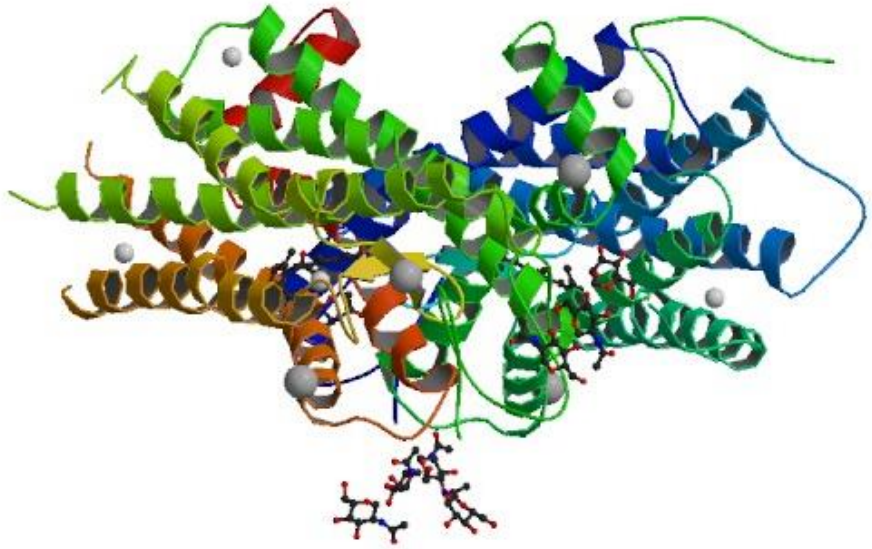
Energy minimization vs simulations



- Energy minimization generates individual minimum energy configurations
- Assuming all minima could be defined, a statistical mechanics partition function would describe the system
  - Possible only for small molecules, small isolated clusters in vacuum (gas)
  - Not feasible for complex systems
- For complex (practically all) systems, computer simulations can be used to probe the potential energy surface

Applications of energy minimization

- Structure optimization
  - Docking studies, structure analysis
  - X-ray structure molecular optimization
- Pre-step before molecular dynamics or Monte Carlo simulations
  - Structural relaxation

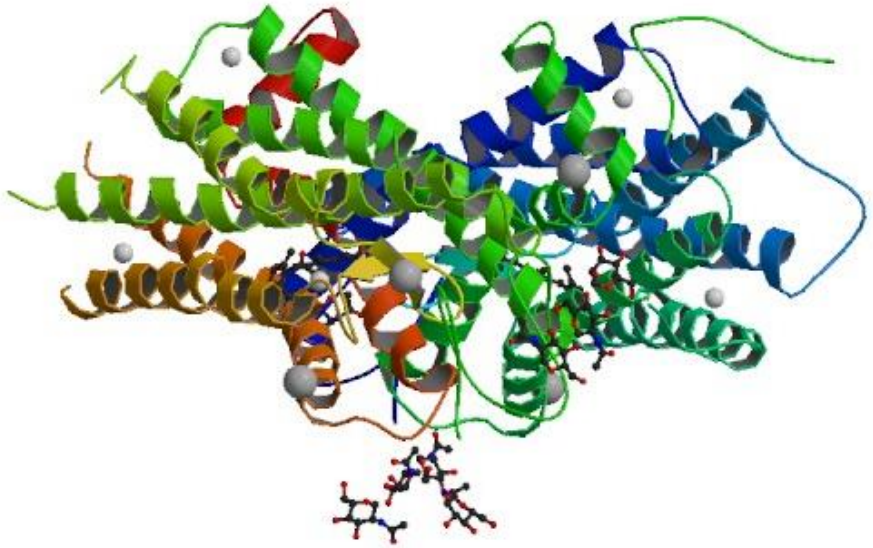


Rhodopsin, PDB entry 1f88

The molecule in the eye that senses light. Composed of a small light-sensitive molecule of retinal, bound inside the protein opsin

Applications of energy minimization

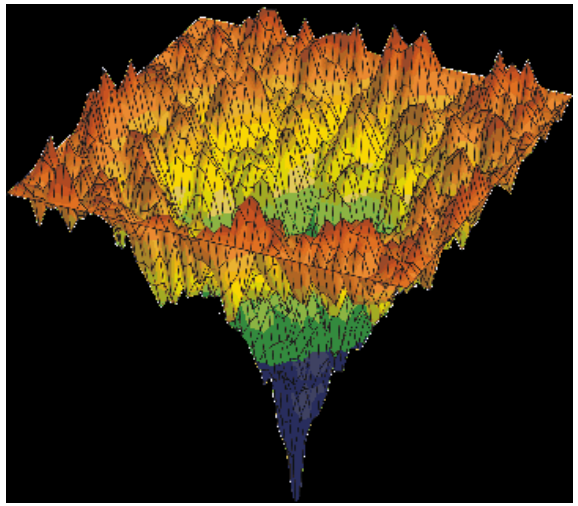
- Normal mode analysis
  - Hessian matrix (=the second-order partial derivatives) provides a force-constant matrix for vibrations in a system
  - Eigenvalues relate to vibrations frequencies
- Transition structures and reaction pathways (saddle points)



The molecule in the eye that senses light. Composed of a small light-sensitive molecule of retinal, bound inside the protein opsin

From molecular conformations to measurable averages

- We have: Potential energy surface
- We need: A measurable quantity
- Now: how do we obtain the measurable quantity from the potential energy surface (force-field)



Time averages, ensemble averages

- Determining experimentally measurable properties of a molecular system requires relating instantaneous values to average measurable value
- At infinite limit

$$A_{ave} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A(p^N(t), r^N(t)) dt$$

Time average

Instantaneous value

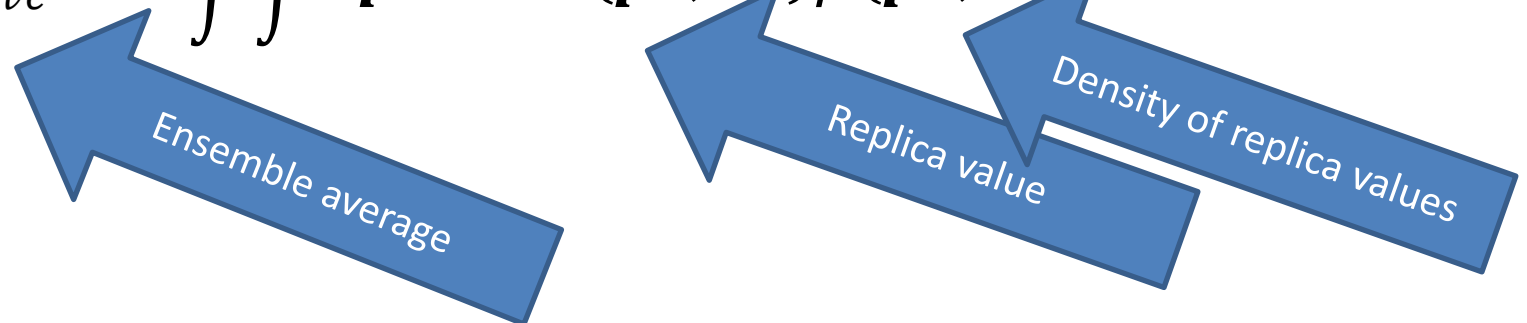
Time averages, ensemble averages

- At infinite time limit, time average A_{ave}

$$A_{\text{ave}} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A(\boldsymbol{p}^N(t), \boldsymbol{r}^N(t)) dt$$

- Calculating this is an issue with any real system (10^{23} atoms)
- Instead, large number of replications: Ensemble average

$$< A_{ave} >= \int \int d\boldsymbol{p}^N d\boldsymbol{r}^N A(\boldsymbol{p}^N, \boldsymbol{r}^N) \rho(\boldsymbol{p}^N, \boldsymbol{r}^N)$$



Ergodic hypothesis

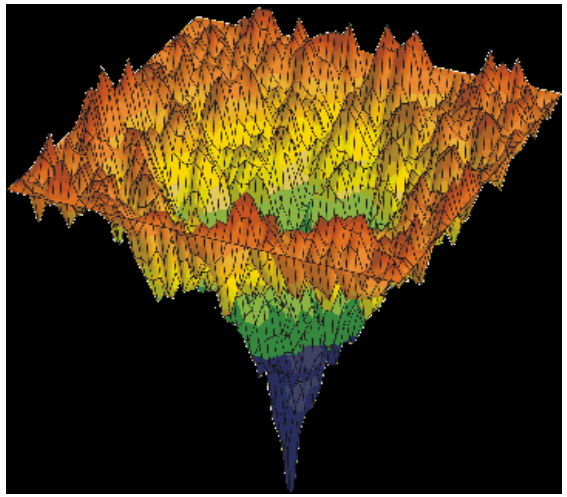
$$A_{\text{ave}} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A(\boldsymbol{p}^N(t), \boldsymbol{r}^N(t)) dt$$

$$< A >= \int \int d\boldsymbol{p}^N d\boldsymbol{r}^N A(\boldsymbol{p}^N, \boldsymbol{r}^N) \rho(\boldsymbol{p}^N, \boldsymbol{r}^N)$$

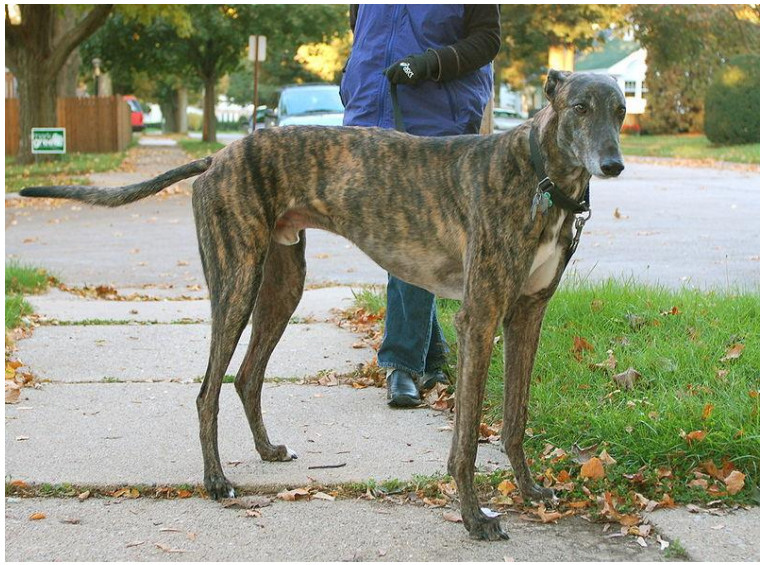
Assume that time average and ensemble average are equal. Valid at ergodic limit.

Key to obtaining “measurable” quantity from molecular simulation

- Model of real system: Potential energy surface
- Measurable quantity
  - ensemble average over a **finite size** ensemble
  - time average over a **finite time**
- If ”finite” would be ”infinite”, no issues, but…

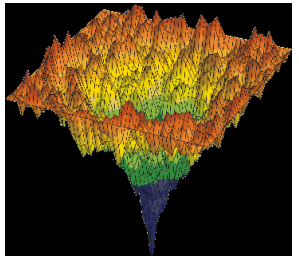


Now to molecular dynamics
Static vs dynamics



vs

Revision:
Basics of molecular dynamics



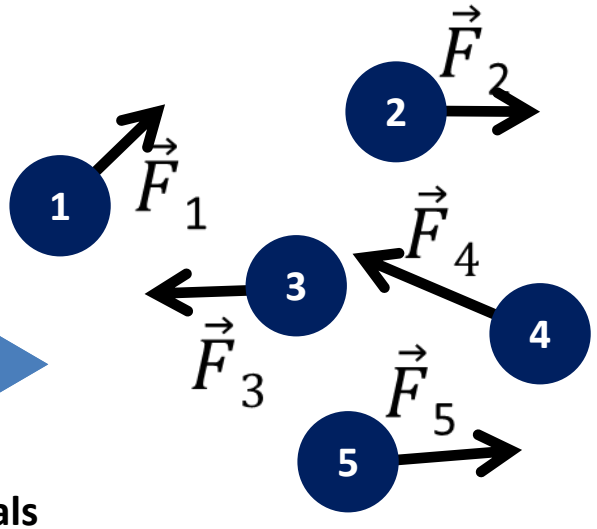
- Potential energy functional E (function of nuclei positions) -> Force on each nuclei

$$\vec{F}=-\nabla E$$

$$\vec{F}=m\vec{a}=m\frac{d\vec{v}}{dt}=m\frac{d^2\vec{r}}{dt^2}$$

time

t_0 $t_0+\delta t$ $t_0+2\delta t$ ...

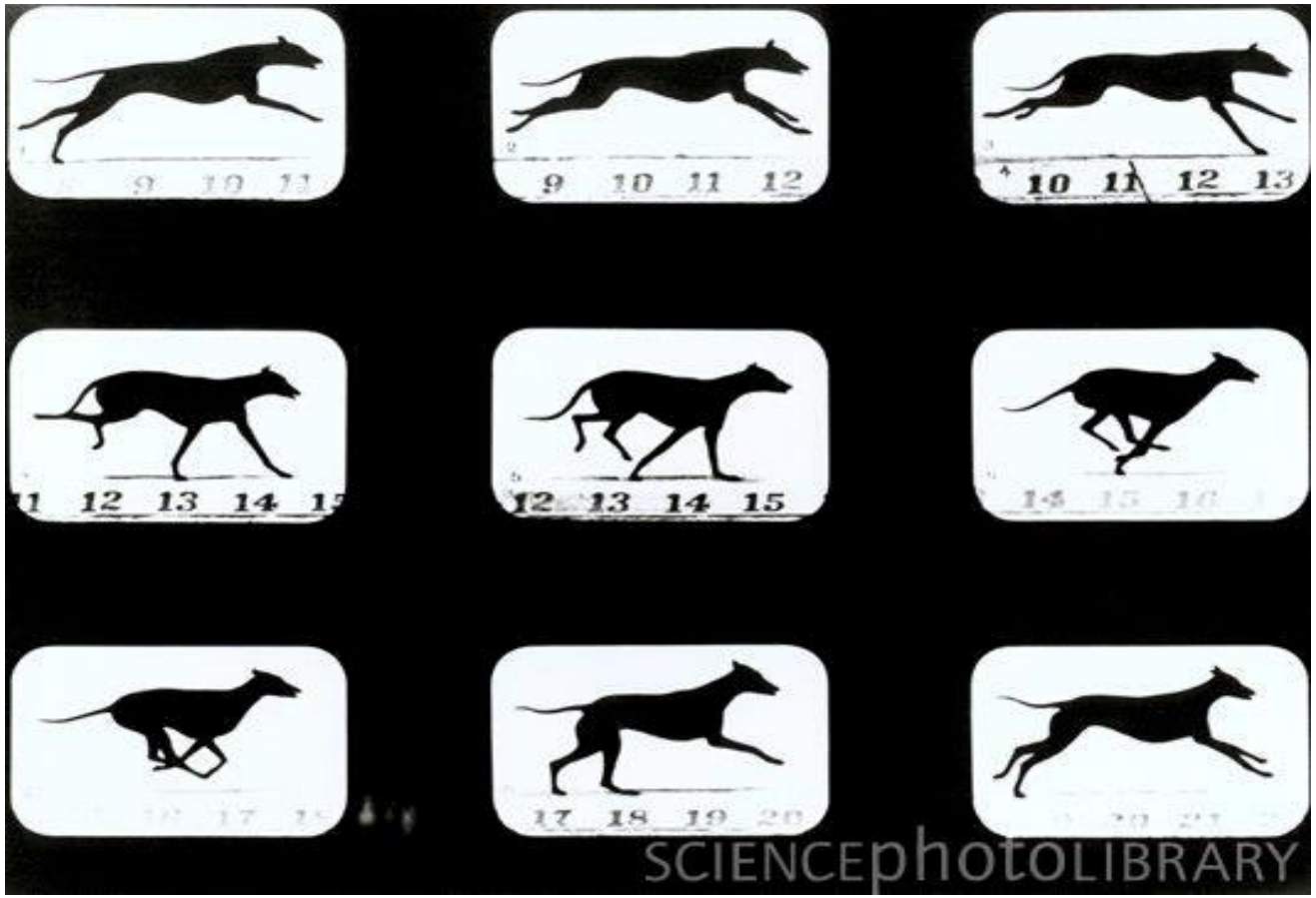


Force for each particle **calculated** at **discrete time intervals**
Particle **positions updated** assuming particle moves with this force (acceleration) in the direction of force for the entire (short) time interval
New forces calculated with updated positions
loop-as-long-as-wanted (typically as long as possible)

Molecular dynamics in brief: sequence of static images

time

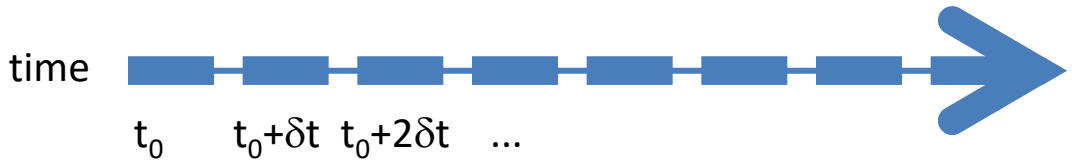
t_0 $t_0+\delta t$ $t_0+2\delta t$ ...



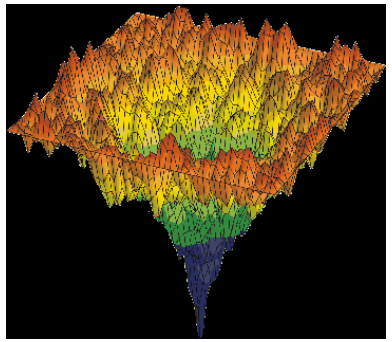
Molecular dynamics

$\vec{F}=-\nabla E$

$\vec{F}=m\vec{a}$



- Any state of the system in future can be predicted from the state right now
  - Deterministic
  - Direct consequence: Any state in the past can be predicted by reversing time in the algorithm
  - Numerical accuracy provides a limit



Molecular dynamics

- Thermodynamic quantities, conformation properties as average corresponding to the configurations that have been present
- M number of time steps

$$A_{\text{ave}} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A(\boldsymbol{p}^N(t), \boldsymbol{r}^N(t)) dt$$

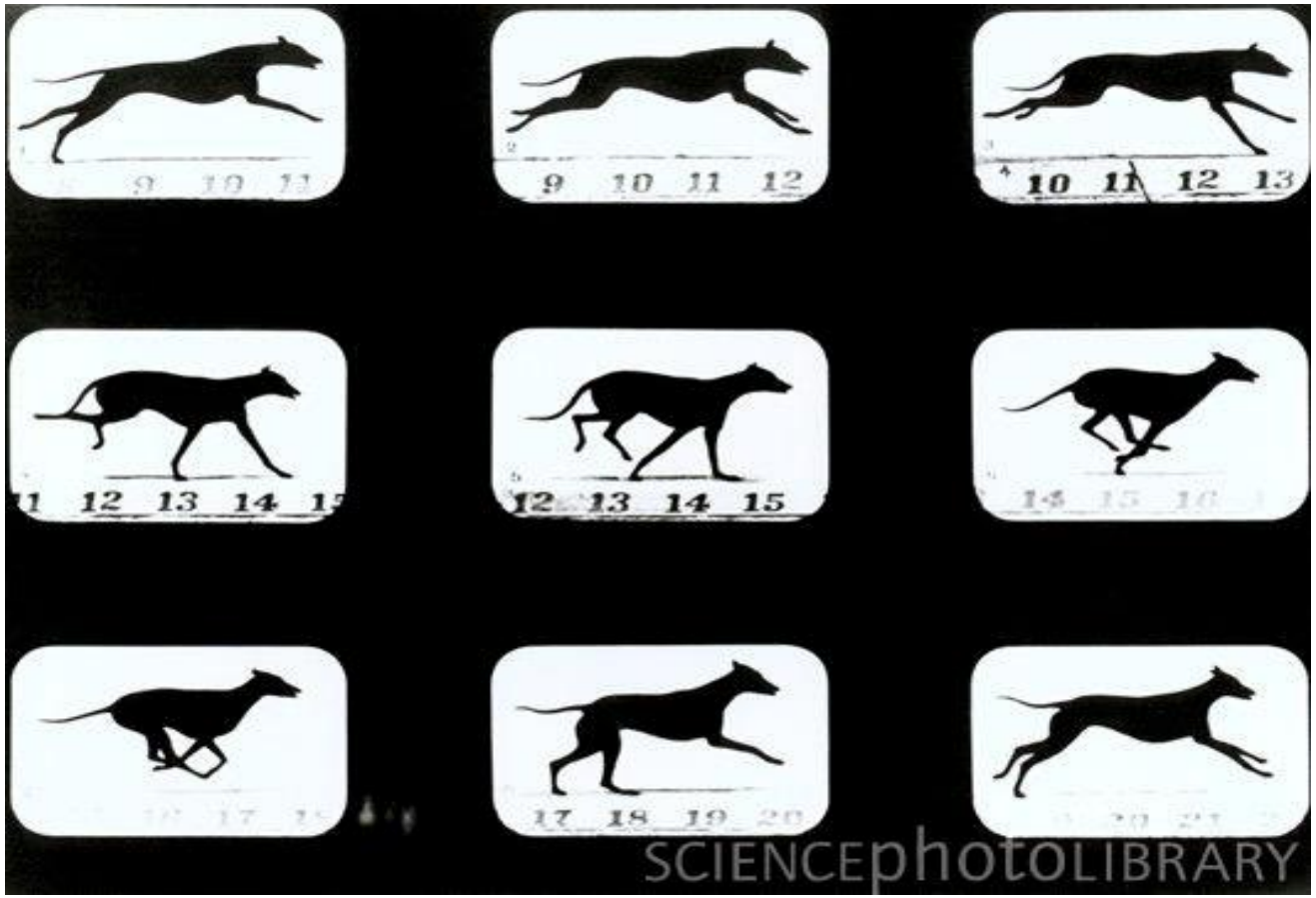
$$< A >= \frac{1}{M} \sum_{i=1}^{M} A(\boldsymbol{r}^N)$$

Molecular dynamics in brief: sequence of static images

$$< A >= \frac{1}{M} \sum_{i=1}^{M} A(\boldsymbol{r}^N)$$

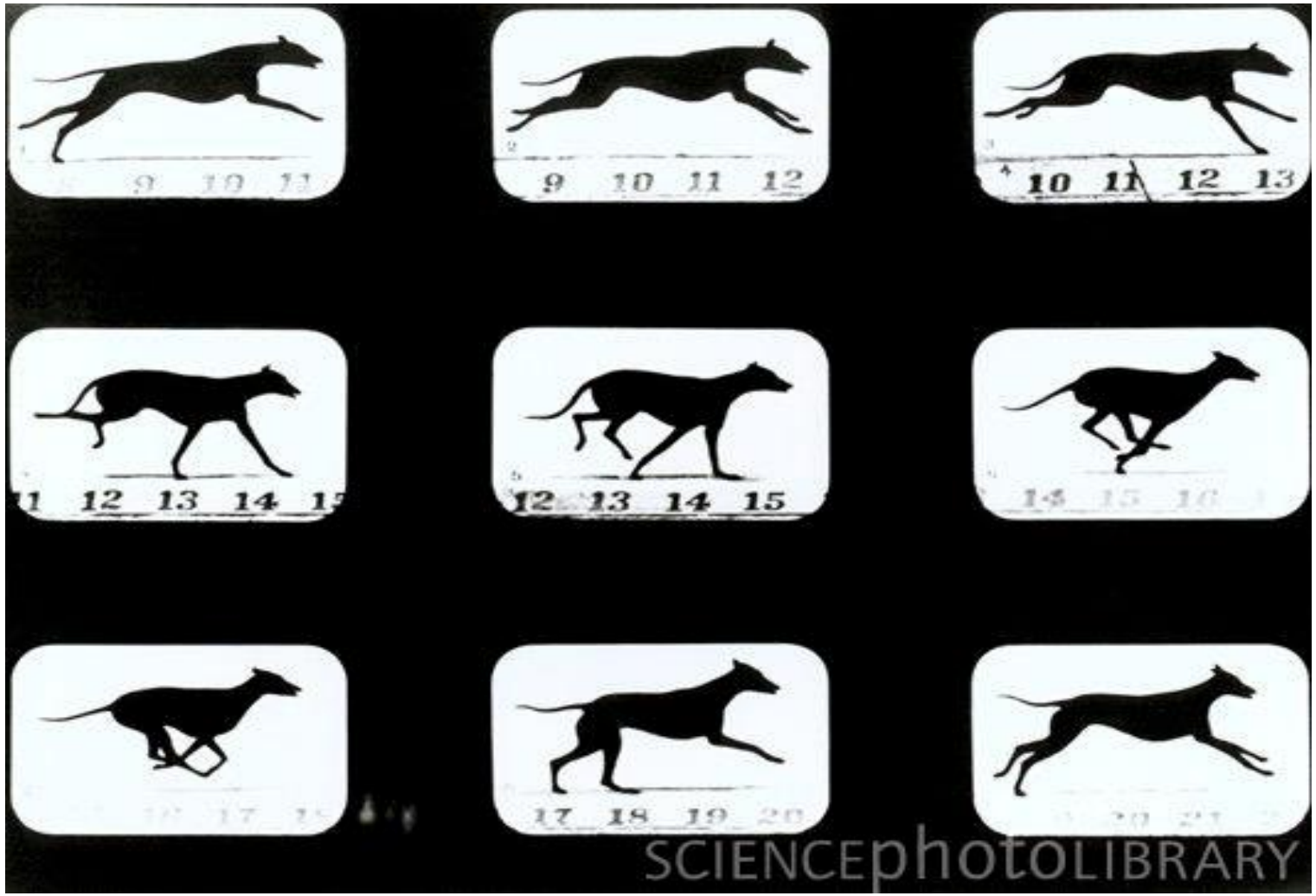
time

t_0 $t_0+\delta t$ $t_0+2\delta t$ ...



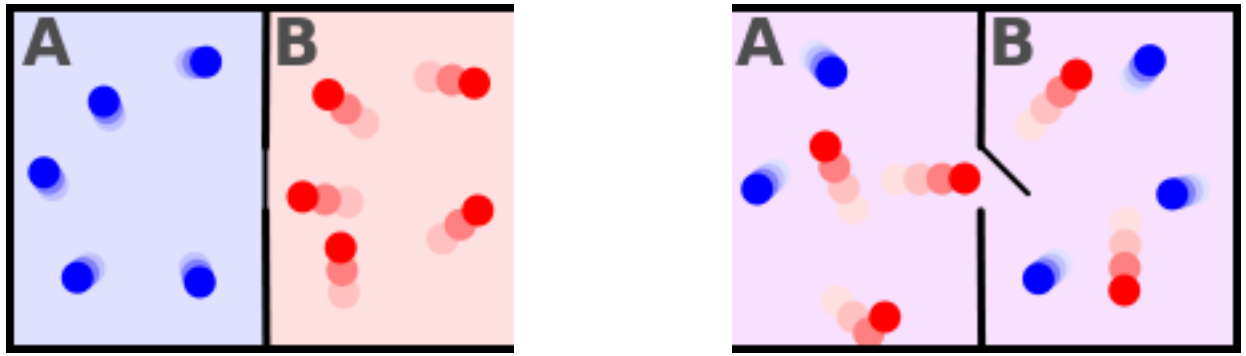
Note: Average may not be representative

$$< A >= \frac{1}{M}\sum_{i=1}^{M} A(\boldsymbol{r}^N)$$



Statistical mechanics ensembles

- Microcanonical ensemble *NVE*
- Canonical ensemble *NVT*
- Isothermal-isobaric ensemble *NTP*
- Grand canonical ensemble μ *VT*



Microcanonical ensemble (*NVE*) is the natural ensemble of molecular dynamics

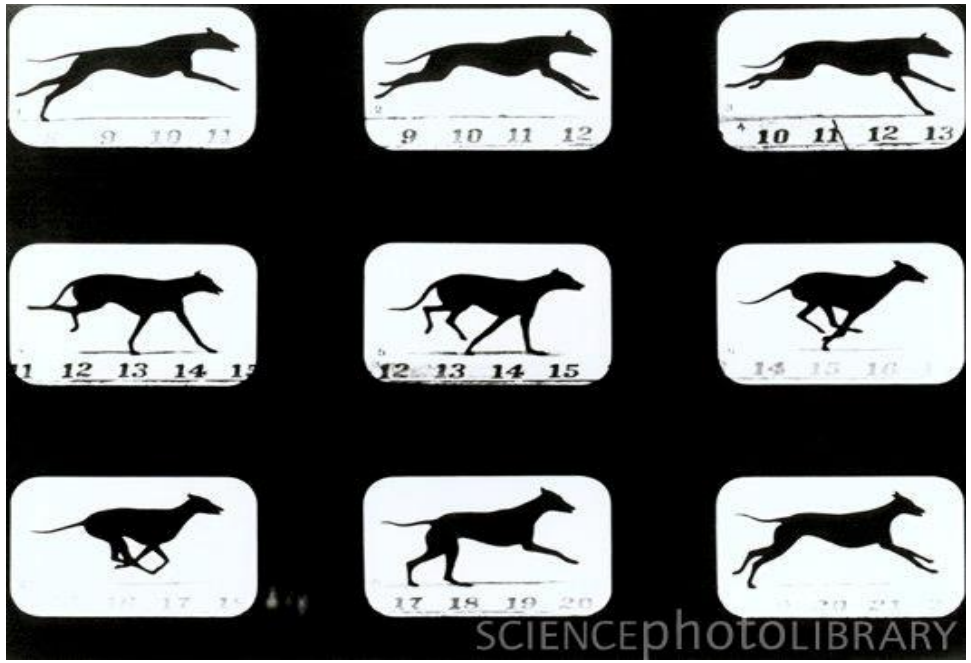
- Number of atoms (*N*), box volume (*V*), and energy (*E*) are conserved
  - *NVE*, microcanonical ensemble
- Equations of motions satisfy naturally energy conservation
- Energy conservation can be used as an inherent check on the implementation
- Free from coupling the microscopic system to macroscopic variables (*NVT* and *NPT* do this)
- In the exercise also *NVT* and *NPT* via algorithm modification (more about this later)

At the end of this lecture, you should know

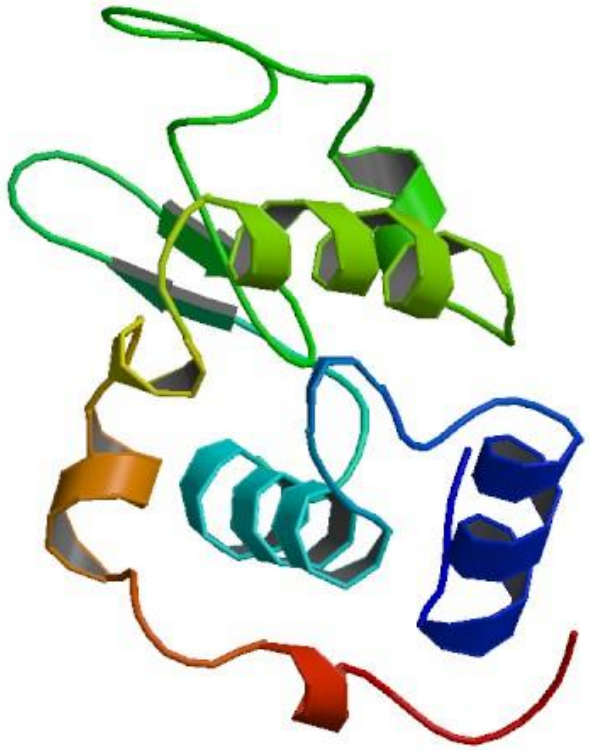
- Connection of potential energy surface and force-field in molecular modelling
  - One defines the other
- Obtaining a measurable quantity that is based on the potential energy surface:
  - Time average and ensemble average
- Concept of molecular dynamics:

  sequence of coordinate and

  velocity snapshots

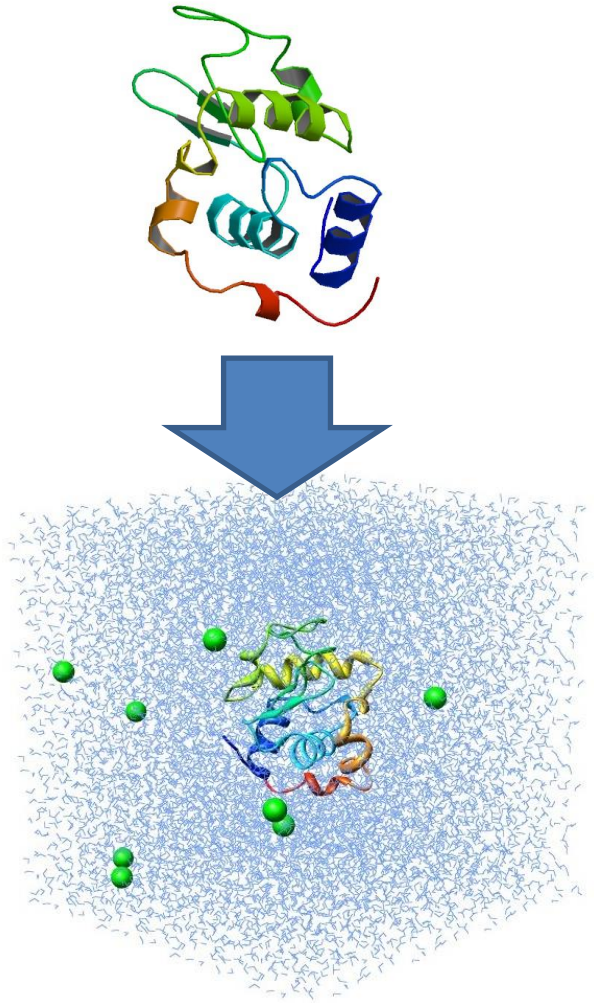


Content of the 1st exercise



- How to find crystal structures of proteins
- Setting up and run a molecular modelling simulation of a small protein
- Analyze the simulation data

How to do a molecular dynamics simulations study for a practical biomolecular system (the small protein)



- Research question?
  - Modelling method according to the relevant length and time scales involved in the phenomena.
  - Design molecule study system so that matches the research question.
  - Appropriate environment (in atomistic detail modelling, for example, solvent such as water and ions or added salt)
- Simulation needs defined
  - “size” -> simulation box
  - “boundaries”
  - Interactions of all atoms / molecules in the system (force-field)
  - Choice of statistical mechanics ensemble (Gibbs free energy / isothermal-isobaric ensemble most common for chemical and biomolecular systems). *T* and *p* controlled by algorithms.
  - System conditions such as molecular concentrations, pressure *p*, temperature *T*, …
  - How is time evolution obtained? Integration algorithm for the equations of motion resulting from forces on each particle.
- Analysis methods / analysis questions

A GROMACS workflow for exercise 1. The workflow takes a PDB (Protein Data Bank) structure file as input and returns a MD trajectory.