COMPUTER SIMULATION OF (BIO)MOLECULAR SYSTEMS

Phys 3170

Prof. Alan E. Mark
Molecular Dynamics, SMMS
a.e.mark@uq.edu.au

SCOPE OF THE LECTURES

1) INTRODUCTION

- what is simulation
- basic choices defining a model
- *choice of the degrees of freedom*
- classical atomistic simulations
- covalent force-field terms
- non-bonded force-field terms
- calculating atomic forces
- force-field parameterization

2) GENERATING CONFIGURATIONS

- searching (incl. energy minimization)
- simulating (molecular dynamics)
- the long-range problem
- spatial boundary conditions
- thermodynamic boundary conditions (temperature, pressure)
- experimentally-derived boundary conditions (X-ray, NMR)]

3) ANALYSIS OF SIMULATIONS

- liquid simulations
- biomolecular simulations
- examples

Literature:

1. Lecture Notes

2. Computer Simulation of Molecular Dynamics: Methodology, Applications and Perspectives in Chemistry
van Gunsteren, W.F. and Berendsen, H.J.C. (1990)
Angew. Chem. Int. Ed. Engl. **29** (1990) 992-1023

Recommended book:

1. Molecular Modelling: Principles and Applications
Andrew R. Leach, Longman Limited 1996

Thursday Lecture

Chemistry Building (68)
8th Floor

Meet on level 3 @ 12:00

Practical exercise

IF LATE WILL HAVE TO PHONE
54180

What you are expected to know.

- Understand some of the applications and limitations of computer modeling.
- Understand the difference between energy minimization and molecular dynamics.
- Name factors that will affect the accuracy and reliability of a simulation.
- Understand the basic components of a force field.
- Be able to explain basic concepts such as:
    - *periodic boundary conditions*
    - *cutoffs*
    - *temperature coupling*
    - *topology*
    - *coordinates*

Why Molecular Modeling

- **Cannot see atoms**

Everything you think you know about proteins is a model.

- **Not everything can be measured**.

Some proteins do not crystallize.

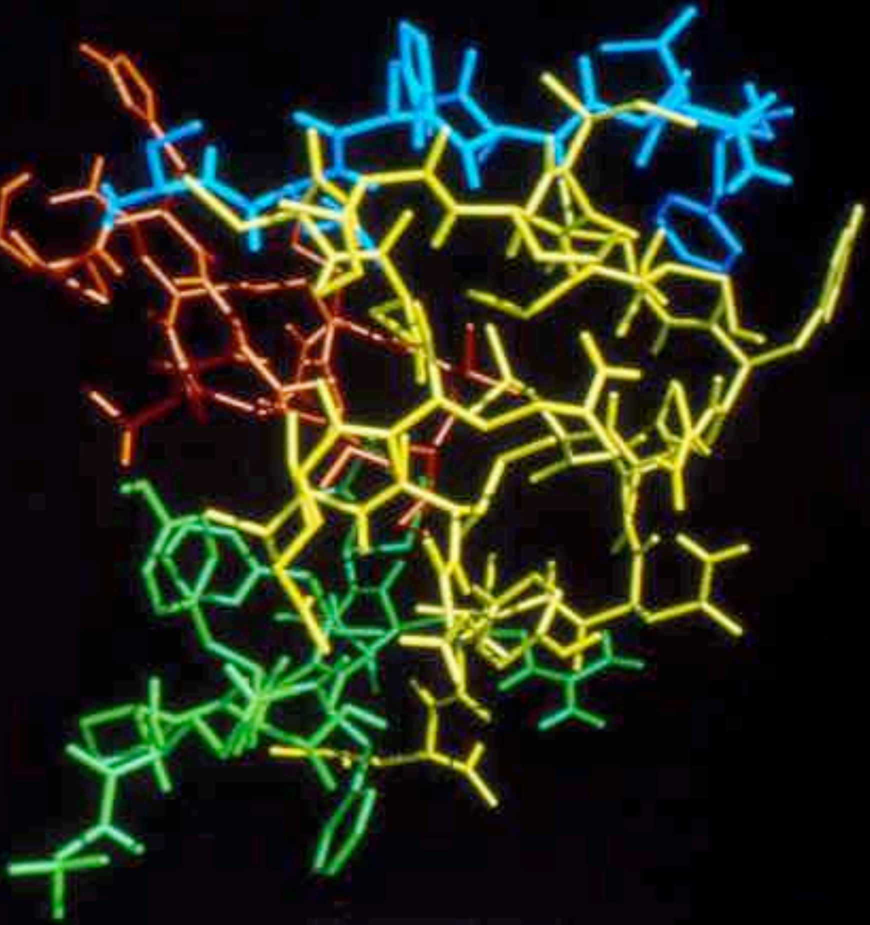
Some are not soluble (NMR).

Some atoms are not observed.

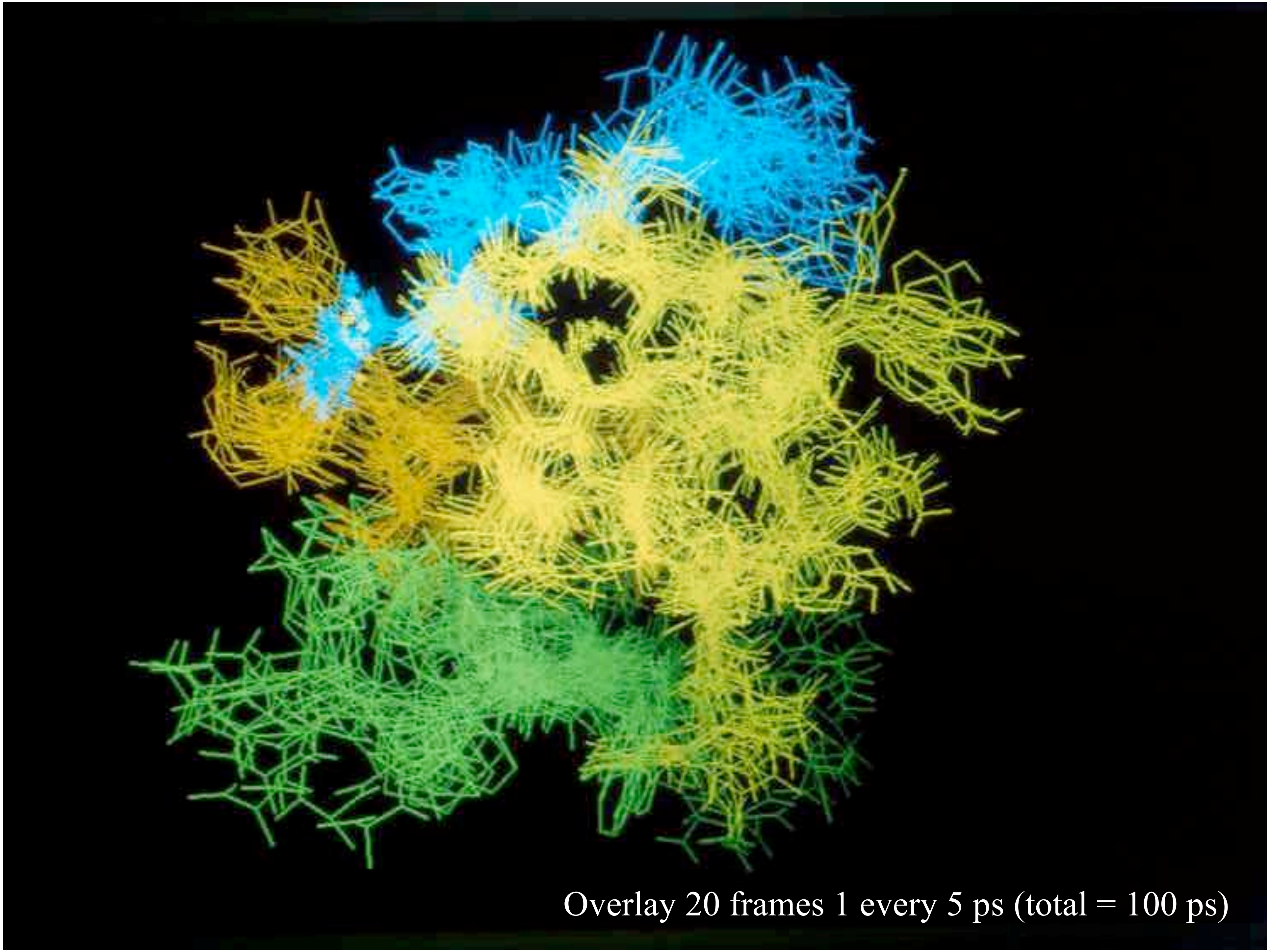
Cannot measure energetics or dynamics at an atomic level.

- **Practical Limitations**

Human genome $> 10^5$ *encoded proteins (PDB* $\sim 10^3$ *3D structures).*

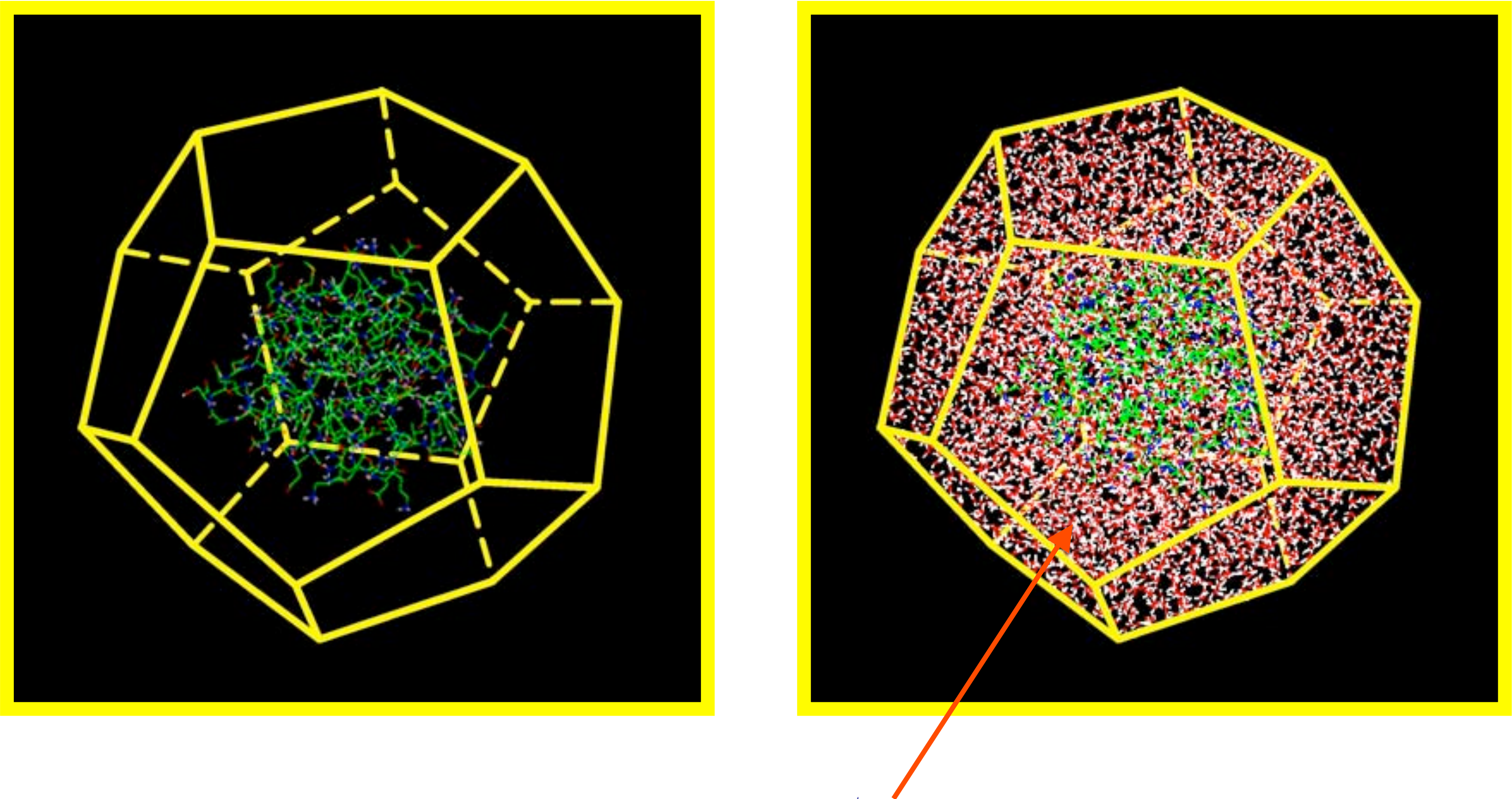


INSULIN



Overlay 20 frames 1 every 5 ps (total = 100 ps)

Periodic Boundary Conditions



water

Molecular Dynamics

A molecular force field describing the inter-atomic interactions

(underlying model)

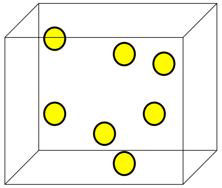
$$
\begin{aligned}
V(r_1, r_2, \ldots, r_N) = {} & \sum_{bonds} \frac{1}{2} K_b (b - b_0)^2 + \sum_{angles} \frac{1}{2} K_\theta (\theta - \theta_0)^2 \\
& + \sum_{torsions} \frac{1}{2} K_\xi (\xi - \xi_0)^2 \\
& + \sum_{torsions} \frac{1}{2} K_\phi [1 + \cos(n\phi + \delta)] \\
& + \sum_{pairs} [C_{12}(i,j) / r_{i,j}^{12} - C_6(i,j) / r_{i,j}^{6}] + q_i q_j / 4\pi\varepsilon_0 \varepsilon r_{i,j}
\end{aligned}
$$

Solve Newton's equations of motion
Time evolution of the system
(classical mechanics)

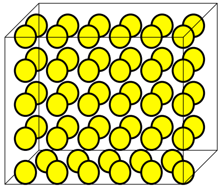
$$\frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i}$$

$$F_i = -\frac{\partial}{\partial r_i} V(r_1, r_2, \ldots, r_N)$$

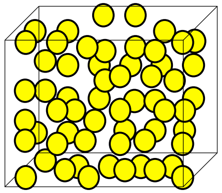
(BIO)CHEMICAL INTEREST



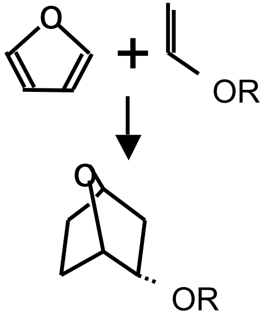
molecules in the gas phase:
-thermodynamic properties
-- molecular structure
- spectroscopic properties



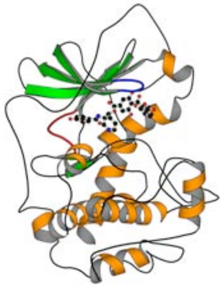
crystals:
- phase behaviour
- molecular structure



liquids/polymers:
- thermodynamics properties
- transport properties
- dielectric properties

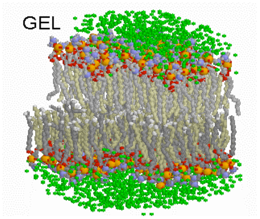


molecules in solution:
- structure and solvation
- conformational equilibria
-reactions

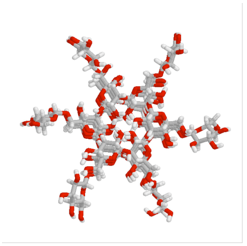


proteins:
- folding, assembly and binding
-dynamics and function
-- effect of mutations

nucleic acids:
- structure and solvation (ions)
- interaction with proteins/ligands
-dynamics



lipids:
- dynamics and phase behaviour
- permeation and diffusion



carbohydrates:
- structure and solvation

When is simulation useful?

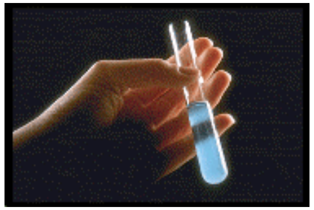
Simulations used instead of experiment when:

- **The process can not be studied experimentally,**
  e.g. *vesicle formation*
- **The process is too expensive to study experimentally,**
- **e.g.** *Structures of small molecules (QM, structures of small peptides in solution)*

Simulations used to complement experiment when:

- **Simulations may reduce the number of experiments to be performed,**
  e.g. *drug design; protein engineering*
- **Simulation reproducing an experiment provides additional insight, e.g.**
  Molecular simulations

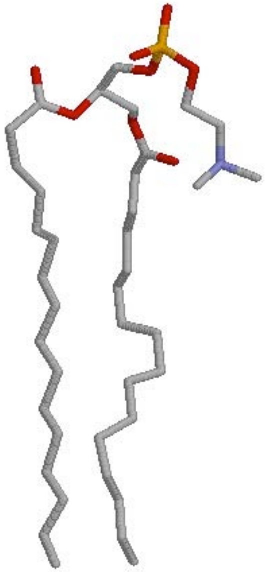
HOW DO MOLECULAR SIMULATIONS PROVIDE INSIGHT ?

	Experiment	*Simulation*
Typical resolution*		
Length :	10^{23} molecules	1 molecule
Time :	1 second	10^{-15} second

*: single molecule / femtosecond is also possible (not simultaneously in condensed phase)

Typical system sizes		
Length :	10^{-3} meter	10^{-9} meter
Time :	10^{3} seconds	10^{-6} second
	low resolution ***large scale***	***high resolution*** ***small scale***

Spontaneous Aggregation of Phospholipid Bilayers.



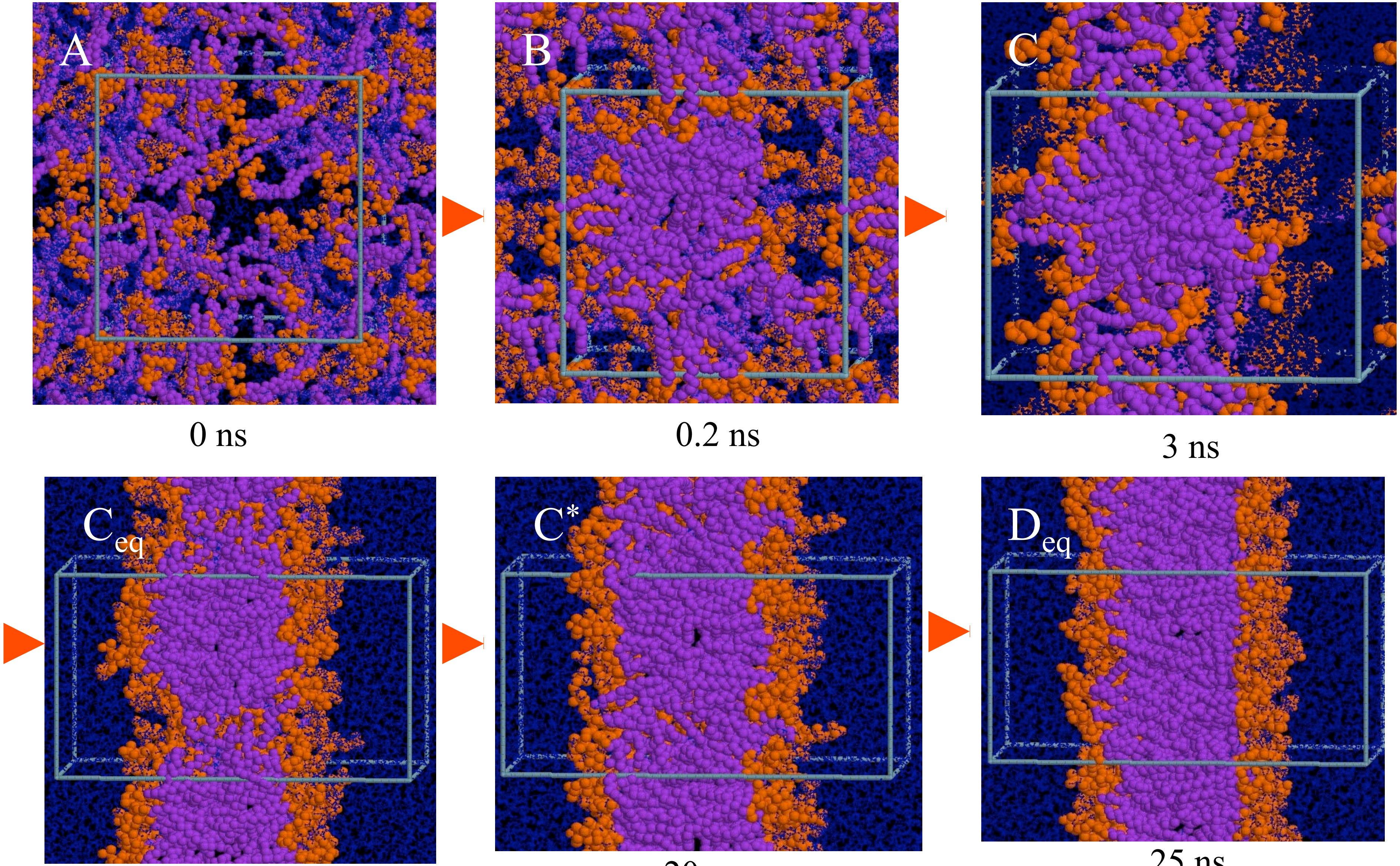
dipalmitoylphosphatidylcholine (DPPC)

- DPPC naturally occurring phospholipid
- well studied computationally

Temp	323 K (fluid phase)
64	DPPC
3000	SPC water

S.J. Marrink, E. Lindahl, O. Edholm and A.E. Mark J. Am. Chem. Soc. **123** (2001) 8638-8639

Spontaneous assembly of phospholipds into a bilayer



0 ns

0.2 ns

3 ns

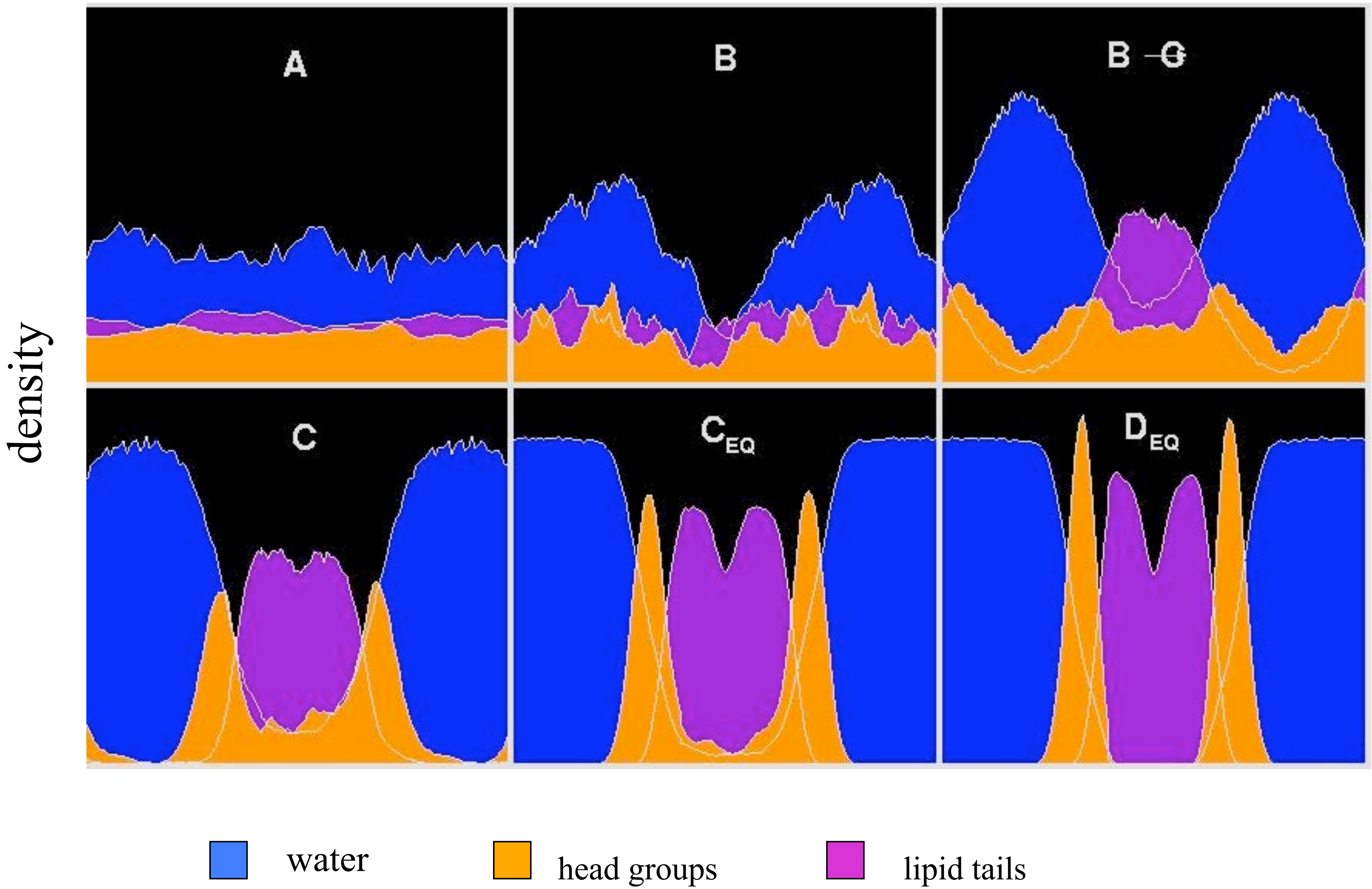
10 ns

20 ns

25 ns

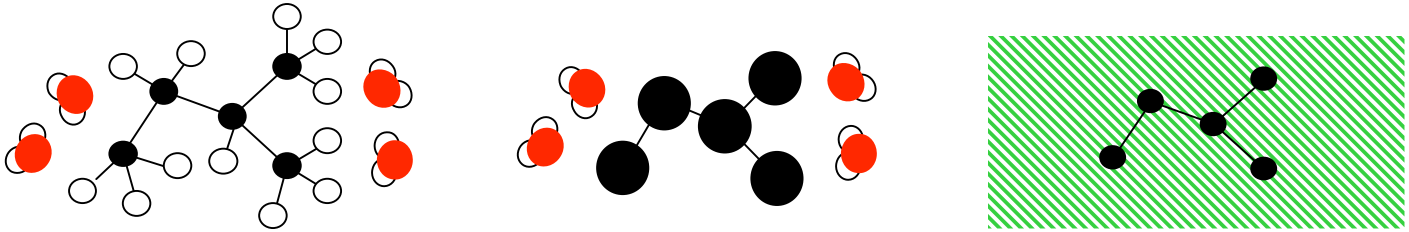
S.J. Marrink

Density Evolution



S.J. Marrink

CLASSICAL ATOMISTIC SIMULATIONS



Ever growing importance in (bio)chemistry because:

Well suited for the study of condensed phase problems (most of chemistry) because:

Can sample extended timescales

thermodynamic properties thorough statistical mechanics
experimentally-accessible timescales (now or near future)

Can treat large systems

(bio)macromolecules in solution

Atomic resolution

realistic dynamics at the atomic level

Complementary to experimental data (stucture at atomic resolution, dynamics)

X-ray diffraction, NMR

Major limitations:

Unable to describe proton and electron transfers, and chemical reactions

need for hybrid methods

Empirical parameters (context dependent)

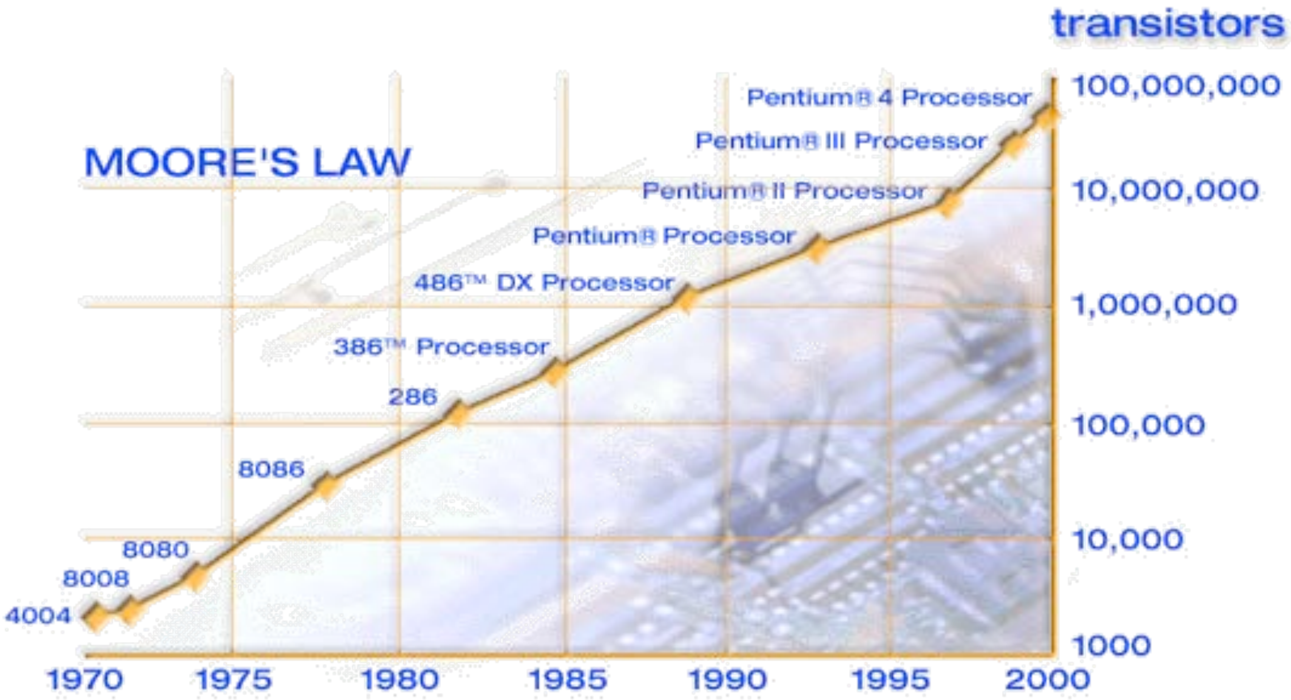
History of classical atomistic simulations

1957: first molecular dynamics simulation (hard disks in 2D)	
1964: Atomic liquid (argon)	10 ps
1971: Molecular liquid (water)	5 ps
1975: Simple short polymer (no solvent)	10 ps
1977: Protein (no solvent)	20 ps
1982: Model membrane (no solvent)	200 ps
1983: Protein in water	20 ps
1986: Nucleic acid in water	100 ps
1989: Protein/nucleic acid complex in water	100 ps
1996: Protein/membrane system in water	100 ps
1996: Enzymatic reaction in water	10 ps
1997: Peptide folding in solution	100 ns
1998: Protein(?) folding(?) in water	1 μs
2000-2001: spontaneous micelle and membrane formation in water	~50 ns
2002: membrane fusion	~100 ns
Current standard: biomolecules in water (10000-100000 atoms)	~10 ns

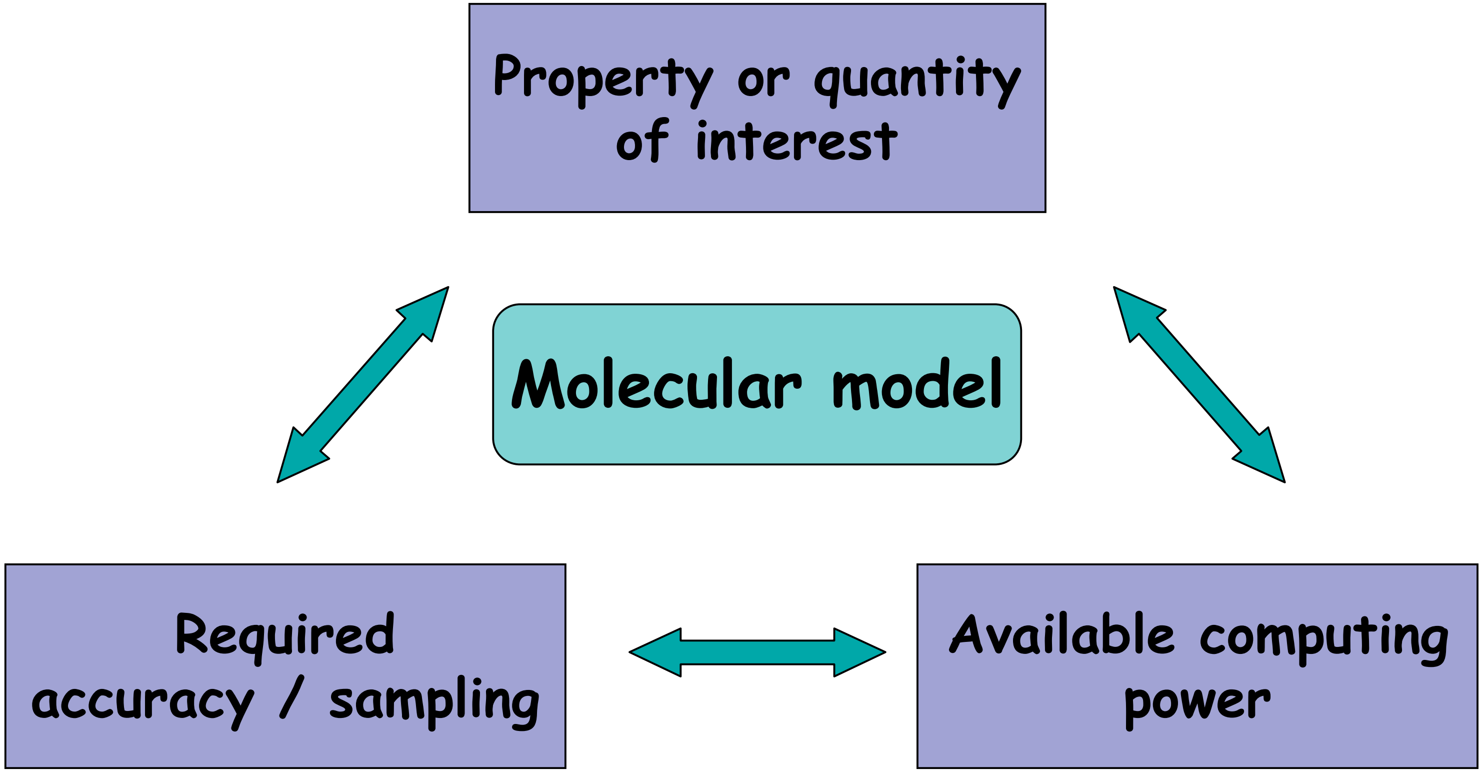
in ~ 10-100 CPU days (10^{14} slower than nature)

Future of classical atomistic simulations

- **Moore's law:**
  - **The computing power increases on average by ~ a factor 10 every 6 year (~factor 3 every 2 years)**



Choice of the model



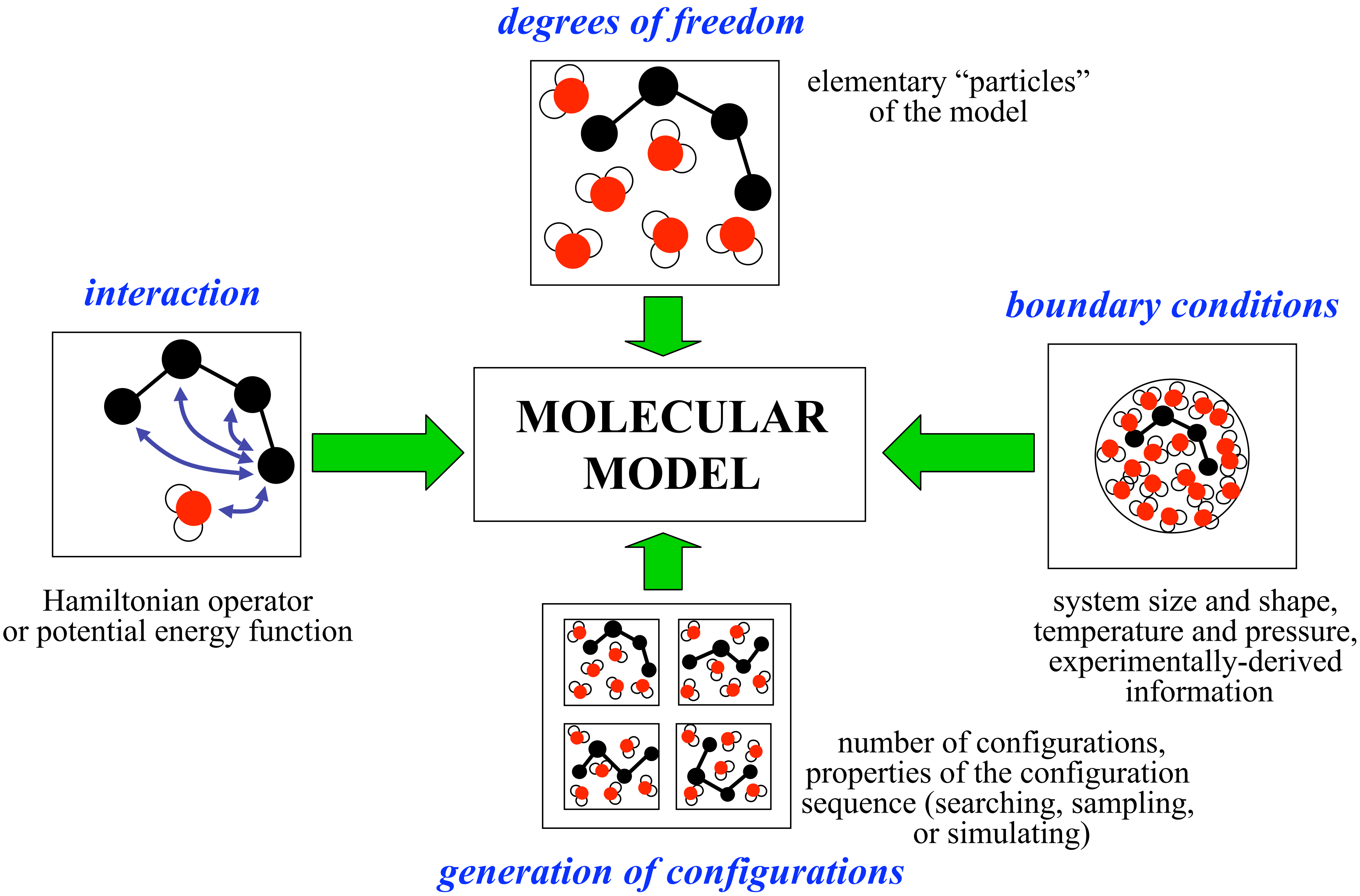
Property or quantity of interest

Molecular model

Required accuracy / sampling

Available computing power

FOUR BASIC CHOICES DEFINING A MOLECULAR MODEL



degrees of freedom

elementary “particles”
of the model

interaction

Hamiltonian operator
or potential energy function

MOLECULAR MODEL

boundary conditions

system size and shape,
temperature and pressure,
experimentally-derived
information

generation of configurations

number of configurations,
properties of the configuration
sequence (searching, sampling,
or simulating)

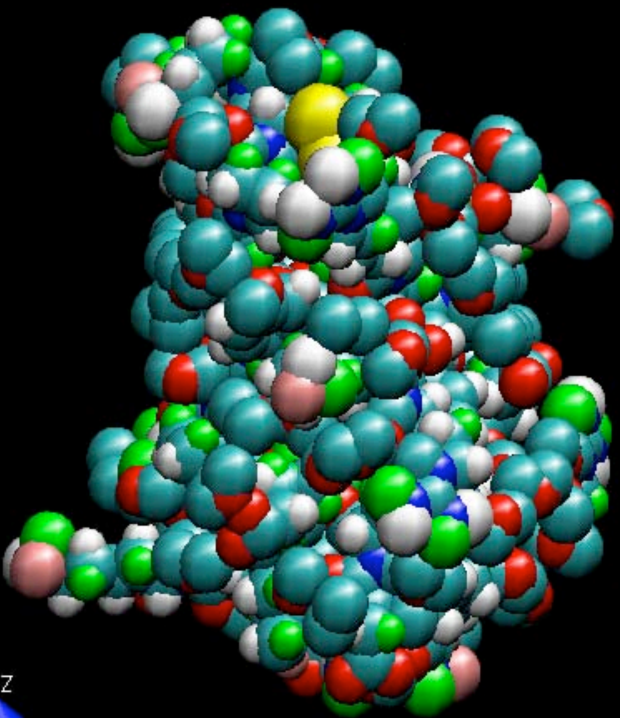
Protein Data Bank File

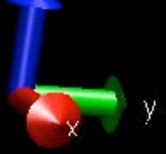
```
HEADER    PROTEINASE INHIBITOR (TRYPSIN)          05   -OCT-84    5PTI      5PTI    3
COMPND    TRYPSIN INHIBITOR (CRYSTAL FORM /II$)                               5PTI    4
SOURCE    BOVINE (BOS $TAURUS) PANCREAS                                       5PTI    5
AUTHOR    A.W LODAWER,R.HUBER                                                 5PTI    6
REMARK   9  15 -JAN-90.                                                       5PTID   5
SEQRES   1     58  ARG PRO ASP PHE CYS LEU GLU PRO PRO TYR THR GLY PRO  5PTI  46
SEQRES   2     58  CYS LYS  ALA ARG ILE ILE ARG TYR PHE TYR ASN ALA LYS  5PTI  47
SEQRES   3     58  ALA GLY LEU CYS GLN THR PHE VAL TYR GLY GLY CYS ARG  5PTI  48
SEQRES   4     58  ALA LYS ARG ASN ASN PHE LYS SER ALA GLU ASP CYS MET  5PTI  49
SEQRES   5     58  ARG THR CYS GLY GLY   ALA                              5PTI  50
HET     PO4    70        5     PHOSPHATE ION                                  5PTI  58
HET     UNK   324        1     UNKNOWN ION PROBABLY A POTASSIUM               5PTI      59
FORMUL   2  PO4     O4 P1  --                                                  5PTI  60
FORMUL   3  DOD    *63(D2 O1)                                                 5PTI  61
SSBOND   1 CYS       5    CYS     55                                          5PTI  67
SSBOND   2  CYS     14     CYS     38                                          5PTI  68
SSBOND   3 CYS      30    CYS     51                                          5PTI  69
CRYST1   74.100   23.400   28.900  90.00  90.00  90.00 P 21 21 21     4  5PTI  70
ATOM      1  N    ARG      1       32.231  15.281  -13.143  1.00 28.28       5PTI  77
ATOM      2  CA   ARG     1       32.184  14.697  -11.772  1.00 27.90       5PTI  78
ATOM      3  C    ARG     1       33.438  13.890  -11.387  1.00 24.90       5PTI  79
ATOM      4  O    ARG     1       34.1  02  13.070 -12.066  1.00 24.44       5PTI  80
ATOM      5  CB   ARG     1       30.797  14.065  -11.625  1.00 27.88       5PTI  81
ATOM      6  CG   ARG     1       30.976  12.589  -11.819  1.00 29.61       5PTI  82
ATOM      7  CD   ARG     1       29.608  12.016  -11.694  1.00 31.91       5PTI  83
ATOM      8  NE   ARG     1       28.942  12.335  -12.945  1.00 33.51       5PTI  84
ATOM      9  CZ   ARG     1       27.670  12.696  -13.050  1.00 34.29       5PTI  85
ATOM     10  NH1 ARG     1       26.901  12.777  -11.999  1.00 3 4.48       5PTI  86
ATOM     11  NH2 ARG     1       27.161  12.963  -14.255  1.00 35.44       5PTI  87
ATOM     12 1D   ARG     1       32.983  14.824  -13.703  1.00 27.71       5PTI  88
ATOM     13 2D   ARG     1       31.275  15.112  -13.535  1.00 28.50       5PT I  89
ATOM     14 3D   ARG     1       32.174  16.346  -13.050  1.00 28.23       5PTI  90
ATOM     15  HA  ARG     1       32.192  15.563  -11.115  1.00 26.97       5PTI  91
ATOM     16 1HB  ARG     1       30.392  14.428  -10.697  1.00 28.71       5PTI  92
ATOM     17 2HB  ARG      1      30.182  14.438  -12.437  1.00 28.97       5PTI  93
ATOM     18 1HG  ARG     1       31.369  12.359  -12.800  1.00 29.44       5PTI  9 4
```

coordinates **x** **y** **z**

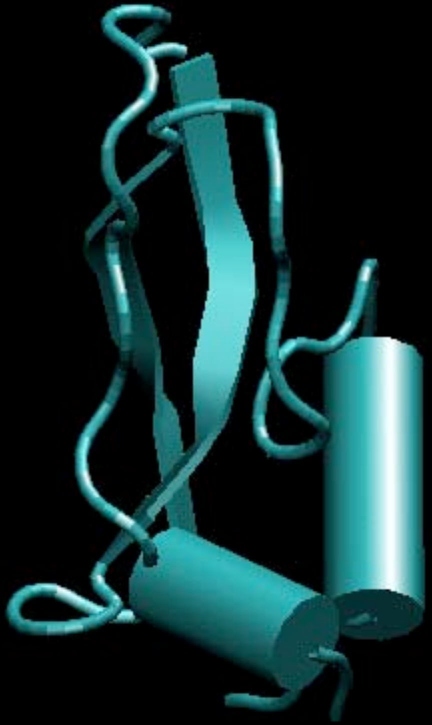
Bovine Pancreatic Trypsin Inhibitor (5PTI)
Different representations

Tube



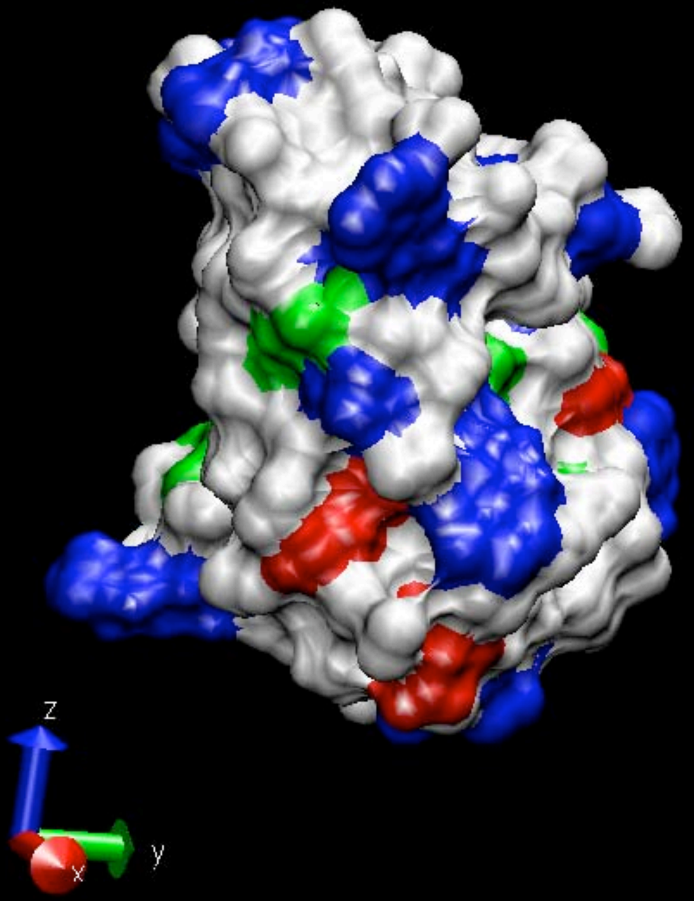


Space Filling
CPK



Cartoon

Bovine Pancreatic Trypsin Inhibitor (5PTI)



How do we represent a bio-molecular system?

Which degrees of freedom to include?

↑ more costly (less can calculate)

Include	Consideration
• electrons and nuclei	→ Quantum Mechanics
• all atoms	→ hydrogen atoms
• united non-polar (CH_1, CH_2, CH_3) atoms	→ polar hydrogens
• united polar groups	→ no hydrogens
• united amino-acid residues	
represent by: 1 particle	
2 particles	→ no atoms
...	
• whole molecule as	
sphere	
rod	→ no internal degrees of freedom
disk	

↓ less physically relevant

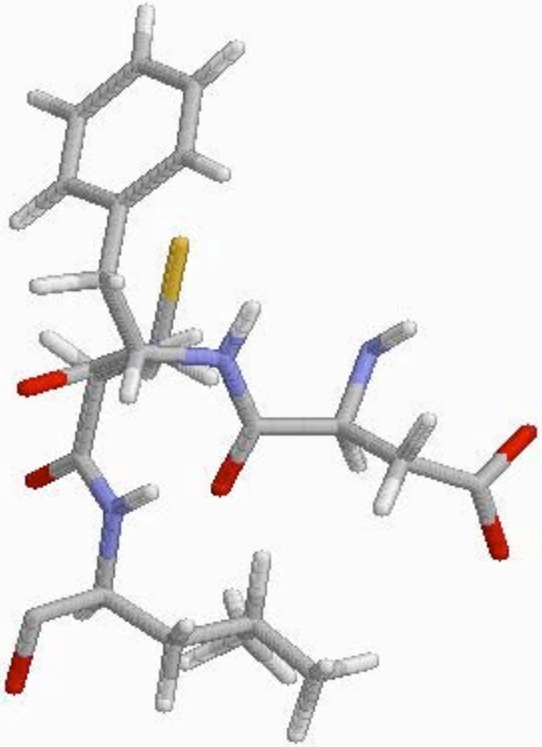
Choice of Degrees of Freedom: Level of Modelling

Asp

Phe

Cys

Leu



What to include?

All atoms

Example: Sites in BPTI (Basic Panceatic Trypsin Inhibitor)

All atoms ~ 1000

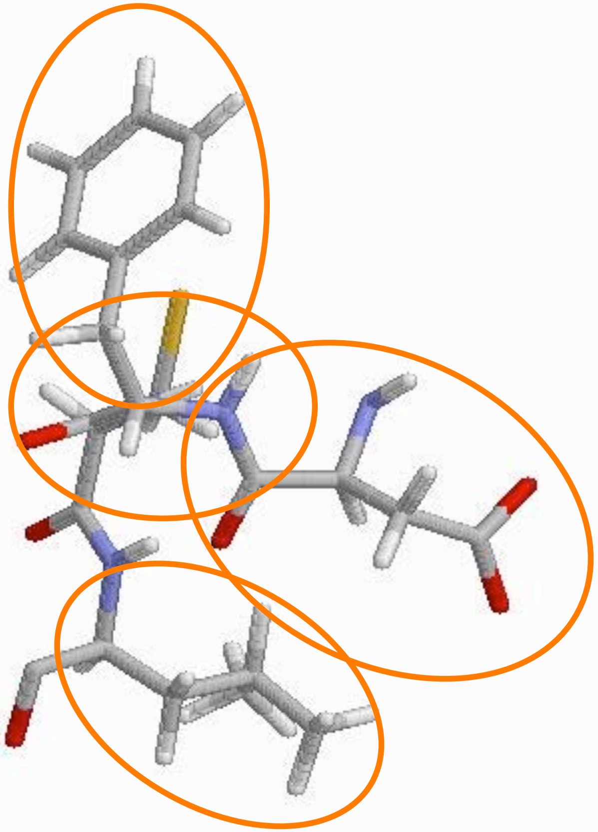
Choice of Degrees of Freedom: Level of Modelling

Asp

Phe

Cys

Leu



What to include?

residues

Example: Sites in BPTI (Basic Panceatic Trypsin Inhibitor)

All atoms	~ 1000
Residues	~ 58

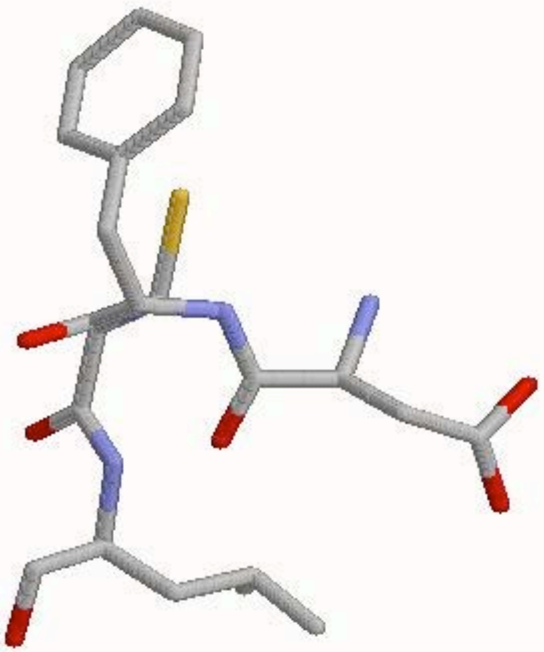
Choice of Degrees of Freedom: Level of Modelling

Asp

Phe

Cys

Leu



What to include?

Heavy atoms

Example: Sites in BPTI (Basic Panceatic Trypsin Inhibitor)
All atoms ~ 1000
Heavy atoms (no hydrogens) ~ 450

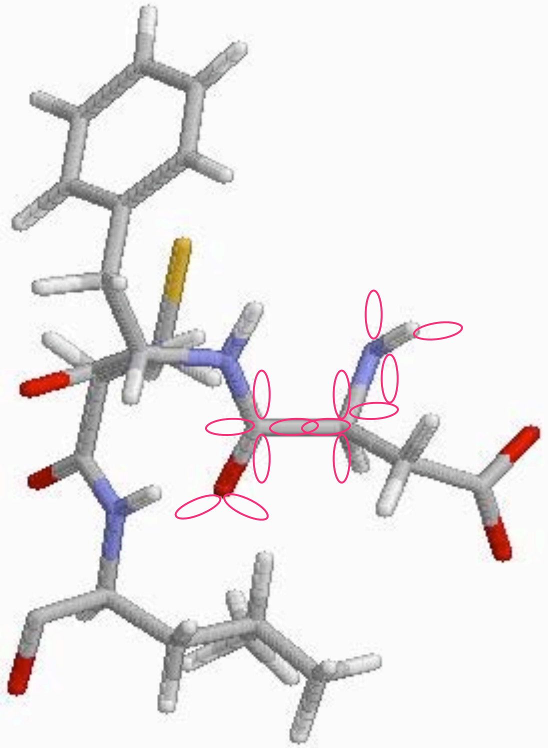
Choice of Degrees of Freedom: Level of Modelling

Asp

Phe

Cys

Leu



What to include?

All atoms

electrons

Example: Sites in BPTI (Basic Panceatic Trypsin Inhibitor)

All atoms ~ 1000

All electrons ~ 3500

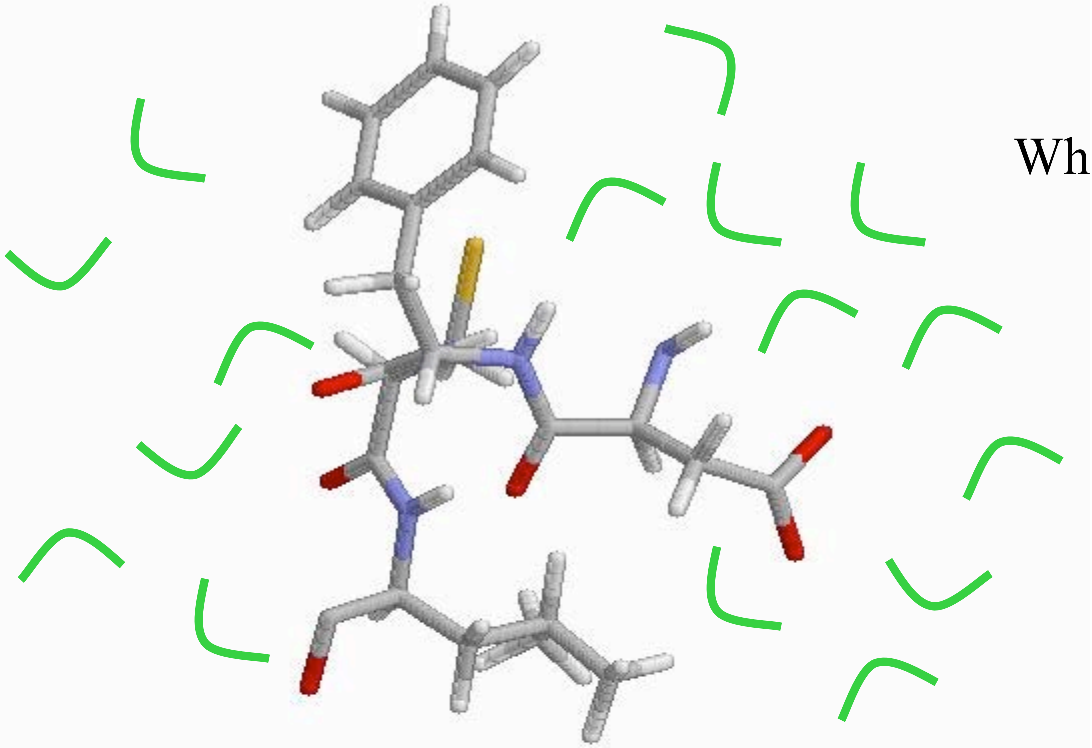
Choice of Degrees of Freedom: Level of Modelling

Asp

Phe

Cys

Leu



What to include?

Explicit Solvent

Example: Sites in BPTI (Basic Panceatic Trypsin Inhibitor)

All atoms	~ 1000
waters	~ 7500

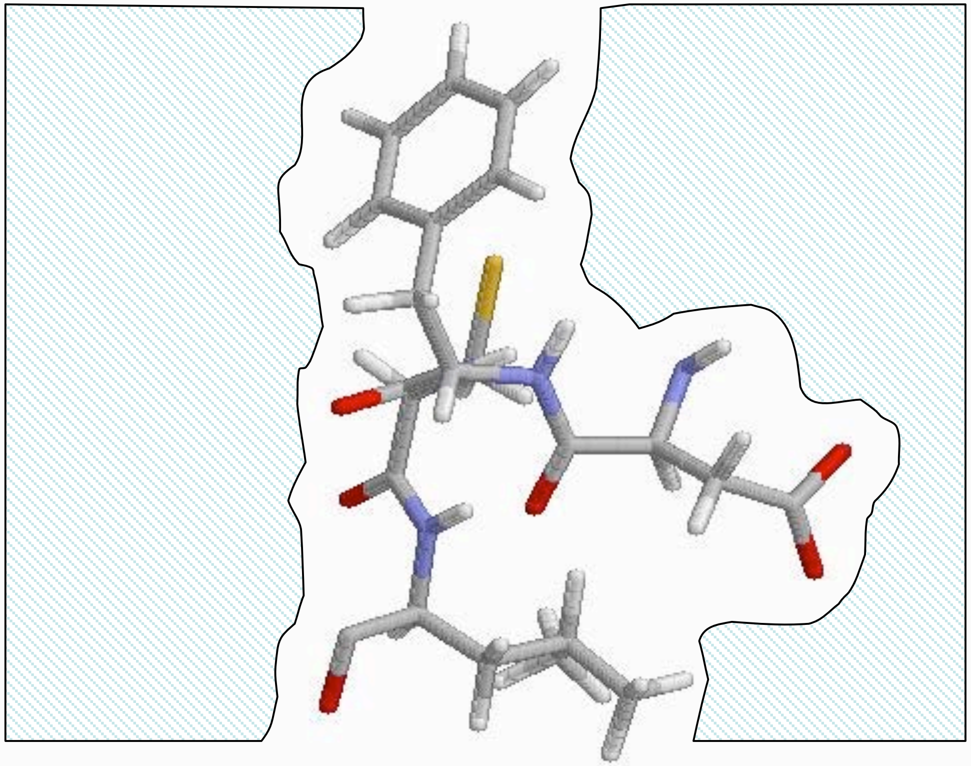
Choice of Degrees of Freedom: Level of Modelling

Asp

Phe

Cys

Leu



What to include?

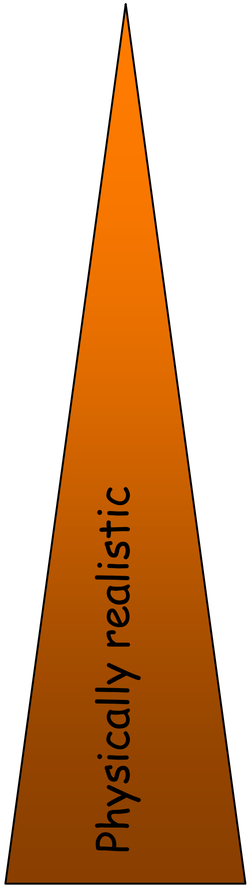
Solvent effects?
- reaction field
- mean force

(+fluctuations)
(+delay)

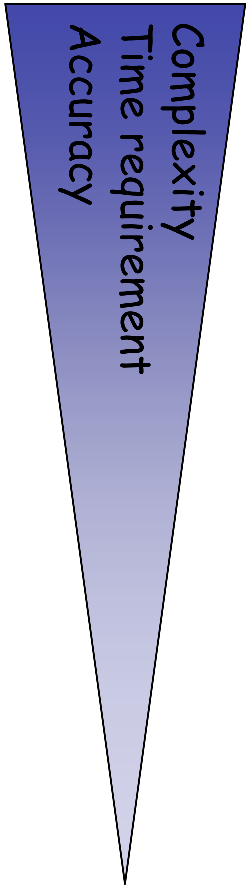
Example: Sites in BPTI (Basic Panceatic Trypsin Inhibitor)

All atoms ~ 1000

Choice of the model

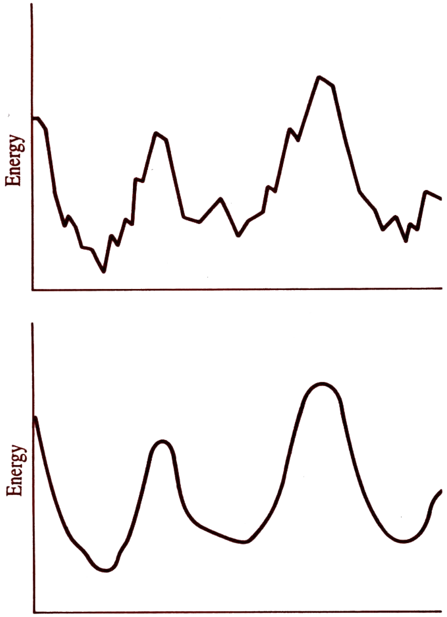


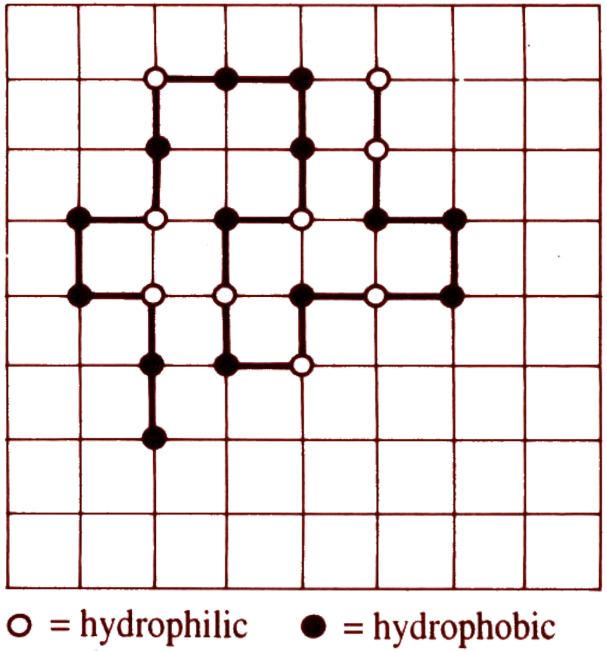
- Treatment of solvent (condensed phase):
  - Explicit:
    - **Large number of particles to deal with**
    - **Computationally expensive**
    - **More realistic energetics and dynamics**
  - Implicit:
    - **Solvent effect is included implicitly (analytical form)**
    - **Less expensive**
    - **Energetics OK, dynamics problematic**
  - No solvent:
    - **Computationally cheap**
    - **Energetics and dynamics problematic**



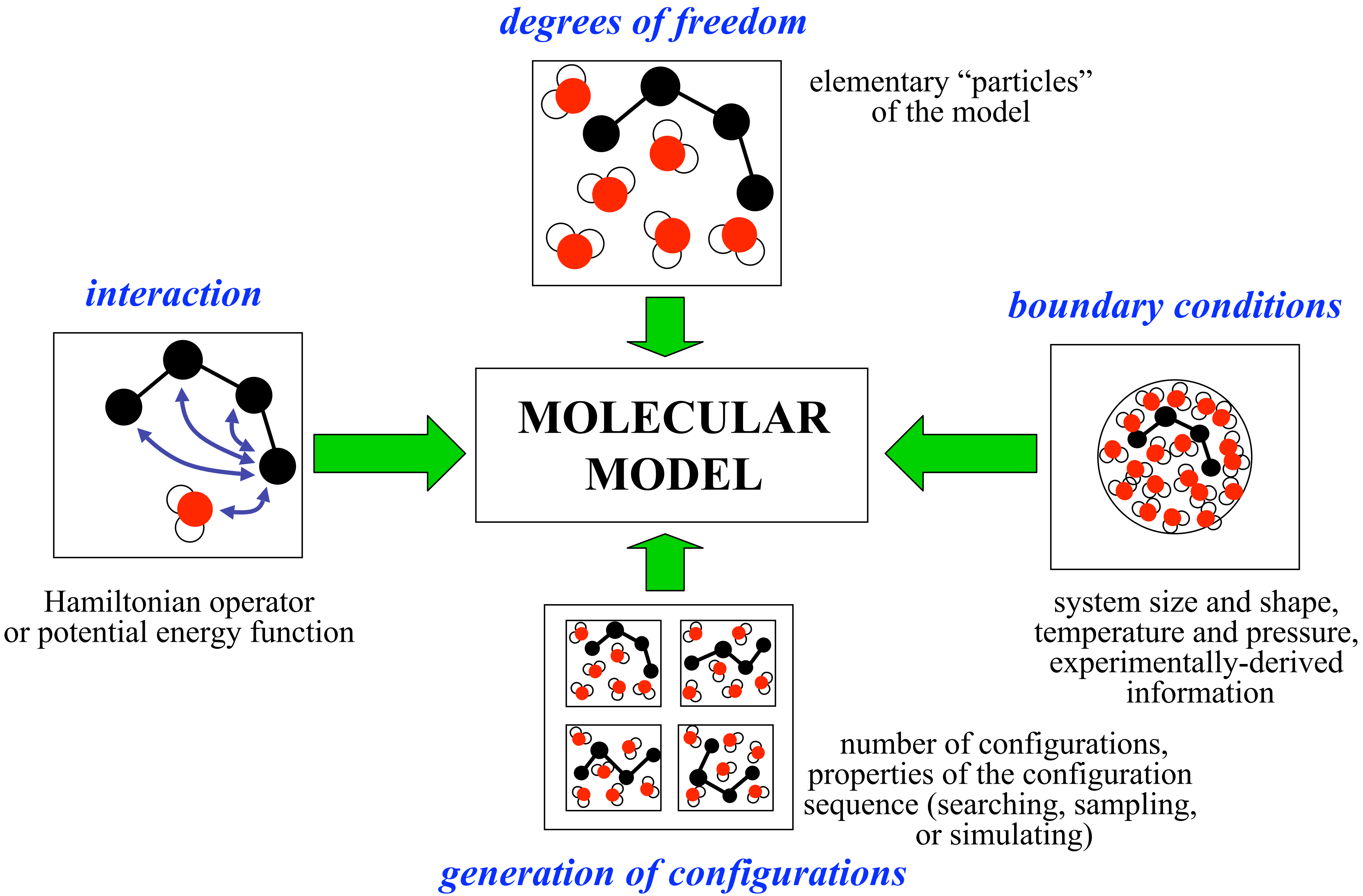
Choice of the model

- Simplifying the model results in simplifying the energy surface (easier search for minima)
- E.g. a simple grid model for studying the mechanism of protein folding





FOUR BASIC CHOICES DEFINING A MOLECULAR MODEL



degrees of freedom

elementary “particles”
of the model

interaction

Hamiltonian operator
or potential energy function

**MOLECULAR
MODEL**

boundary conditions

system size and shape,
temperature and pressure,
experimentally-derived
information

number of configurations,
properties of the configuration
sequence (searching, sampling,
or simulating)

generation of configurations

Some Basic Definitions (Classical Mechanics)

System of N particles $\vec{r}_1, \vec{r}_2, \cdots, \vec{r}_N$

Cartesian coordinates $\{\vec{r}_1, \vec{r}_2, \cdots, \vec{r}_N\} \equiv \vec{r}^{\,N} \equiv \vec{r}$

Interaction function or energy $V(\vec{r}_1, \vec{r}_2, \cdots, \vec{r}_N)$

Force on particle, *i*

$$\vec{F}_i = \vec{\nabla}_i V \equiv \frac{\partial V}{\partial \vec{r}_i} \equiv \left(\frac{\partial V}{\partial x_i}, \frac{\partial V}{\partial y_i}, \frac{\partial V}{\partial z_i} \right)$$

mass $m_1, m_2, \cdots, m_N$

Velocity $\vec{v}_i \equiv \frac{d}{dt}\vec{r}_i \equiv \dot{\vec{r}}_i$

Acceleration $\vec{a}_i \equiv \frac{d}{dt}\vec{v}_i \equiv \ddot{\vec{r}}_i$

Momentum $\vec{p}_i = m_i \vec{v}_i$

Newton's equations of motion $\vec{F}_i = m_i \ddot{\vec{r}}_i \Leftrightarrow m_i \dot{\vec{v}}_i = \vec{F}_i ; \dot{\vec{r}}_i = \vec{v}_i$

Kinetic energy

$$K = \sum_{i=1}^{N} \frac{1}{2} m_i \dot{\vec{r}}^{\,2}{}_i = \sum_{i=1}^{N} \frac{\vec{p}^{\,2}}{2m_i}$$

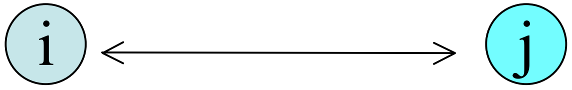
$$= \sum_{i=1}^{N} \frac{1}{2} m_i \left[\dot{x}_i + \dot{y}_i + \dot{z}_i \right]$$

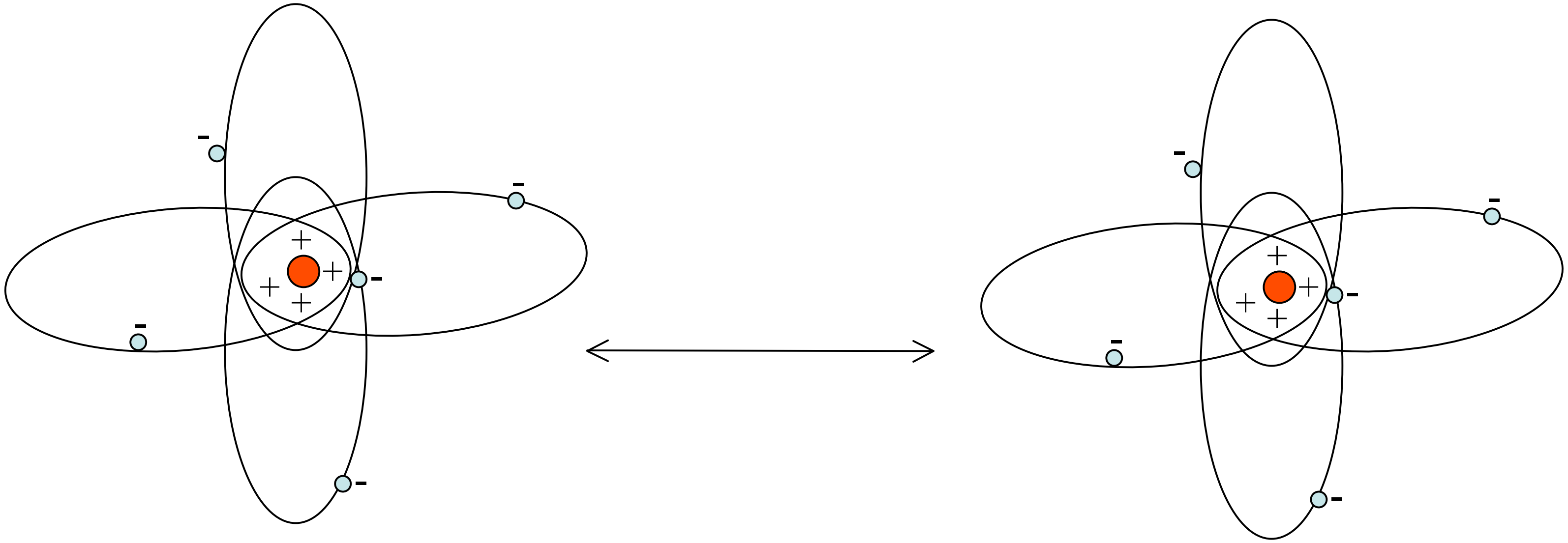
Potential energy $V(\vec{r}_1, \vec{r}_2, \cdots, \vec{r}_N)$

In principle all interactions between atoms are electrostatic and ideally we would describe the system quantum mechanically .

Given this is not possible what is the simplest and computationally most efficient means to approximate the interatomic interactions

$$V_e \propto \frac{Z_i Z_j}{r_{ij}}$$



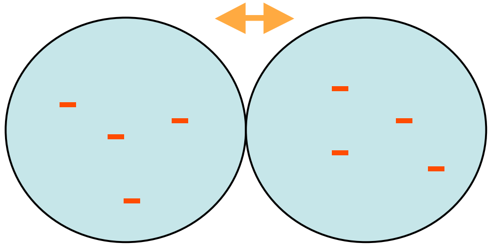


van der Waals non-bonded interactions

To avoid treating all electrons explicitly use effective pair potentials

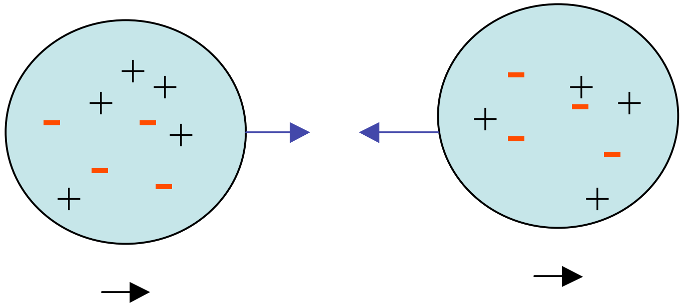
1. van der Waals interactions (interactions between atoms independent of net charge).

A. short range repulsion due to electron overlap



exponential

B. medium range attraction due to dispersion (induced dipoles)

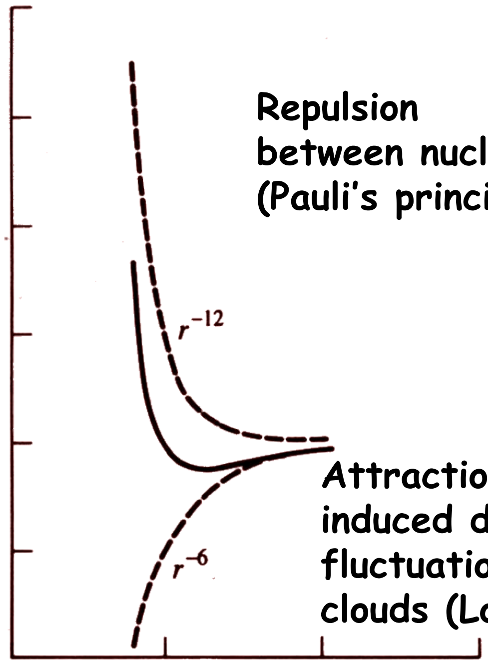


$$V_R = \frac{3I\alpha_i\alpha_j}{4r_{ij}}$$

Van der Waals interaction

- Attractive long-range forces
- Repulsive short-range forces

- Often approximated using the Lennard-Jones 12-6 function

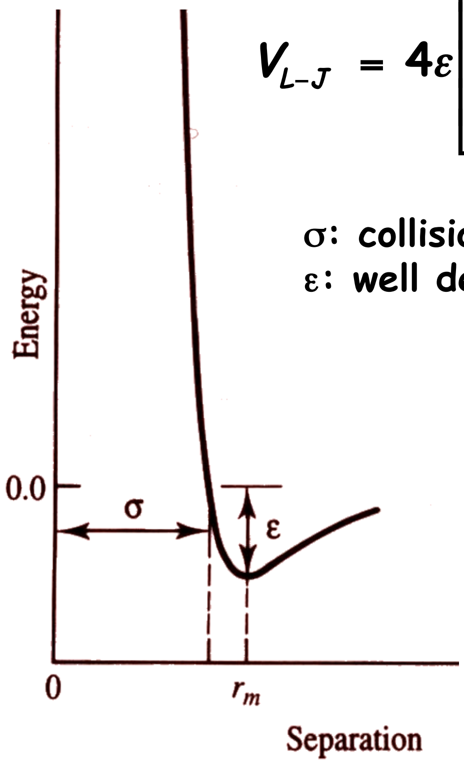


Repulsion between nuclei (Pauli's principle)

Attraction between induced dipoles from fluctuations in electron clouds (London forces)

$$V_{L-J} = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

σ: collision diameter
ε: well depth



Lennard-Jones (6-12) potential

(pairwise force 2-body term)

Energy:

$$V(r)_{vdw} = \frac{A}{r^{12}} - \frac{B}{r^{6}}$$

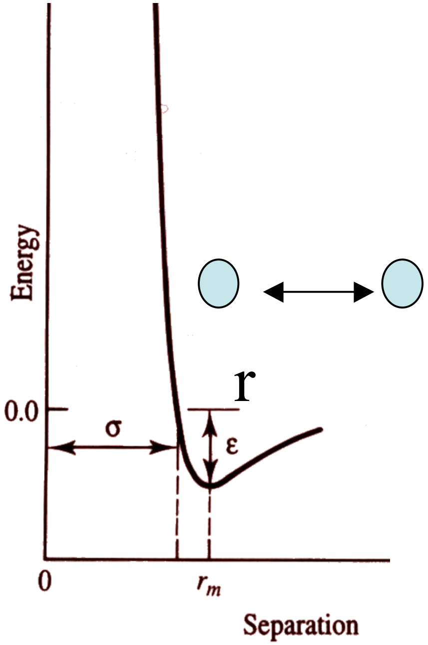
or

$$V(r)_{vdw} = 4\varepsilon\left[\frac{\sigma^{12}}{r^{12}} - \frac{\sigma^{6}}{r^{6}}\right]$$

where

$$\vec{r}_{ij} = \vec{r}_i + \vec{r}_j$$

$$\vec{r}_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$$



Force: (need x, y and z components)

$$F_{x_i} = \frac{d}{dx_i} V(r_1, r_2, \ldots, r_N)$$

Derivatives of the energy function

- Determination of the force requires the derivative of the energy to be calculated with respect to the coordinates.

The force on atom *i* due to *V(r_{ij})* is given by:

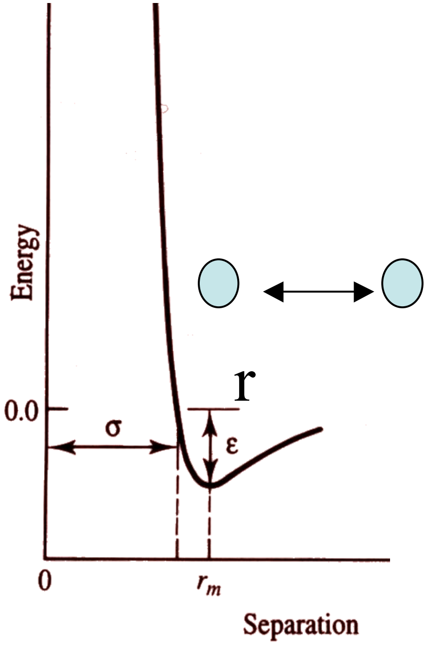
$$F_i(r_{ij}) = -\frac{\partial V(r_{ij})}{\partial x_i}$$

•The derivative can be calculated using the chain rule:

$$\frac{\partial V(r_{ij})}{\partial x_i} = \frac{\partial V(r_{ij})}{\partial r_{ij}} \times \frac{\partial r_{ij}}{\partial x_i}$$

Lennard-Jones (6-12) potential

(pairwise force 2-body term)



Energy:

$$V(r)_{vdw} = \frac{A}{r^{12}} - \frac{B}{r^{6}}$$

or

$$V(r)_{vdw} = 4\varepsilon\left[\frac{\sigma^{12}}{r^{12}} - \frac{\sigma^{6}}{r^{6}}\right]$$

where

$$\vec{r}_{ij} = \vec{r}_i + \vec{r}_j$$

$$\vec{r}_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$$

Force:

$$F_{x_i} = \frac{d}{dx_i} V(r_1, r_2, \ldots, r_N)$$

$$= -\sum_{i \neq j}^{N} \frac{\partial}{\partial r_i} V(r_{ij}) . \frac{\partial r_{ij}}{\partial x_i}$$

$$\frac{\partial r_{ij}}{\partial x_i} = \frac{\partial}{\partial x_i}\left[(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2\right]^{1/2}$$

$$= \frac{1}{2}[\ldots]^{/2} 2(x_i - x_j)$$

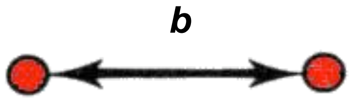
$$= r_{ij}^{\;1/2} x_{ij} = \frac{x_{ij}}{r_{ij}}$$

$$\frac{\partial}{\partial r_i} V(r_{ij}) = 4\varepsilon\left[-12\left(\frac{\sigma}{r_{ij}}\right)^{12} + 6\left(\frac{\sigma}{r_{ij}}\right)^{6}\right]\frac{1}{r_{ij}}$$

$$\boxed{F_{x_i} = 4\varepsilon\left[-12\left(\frac{\sigma}{r_{ij}}\right)^{12} + 6\left(\frac{\sigma}{r_{ij}}\right)^{6}\right]\frac{x_{ij}}{r_{ij}^{\;2}}}$$

Atomic Interaction Functions

Covalent Bond interaction (2 body term)



Various forms possible, e.g.:

Morse potential

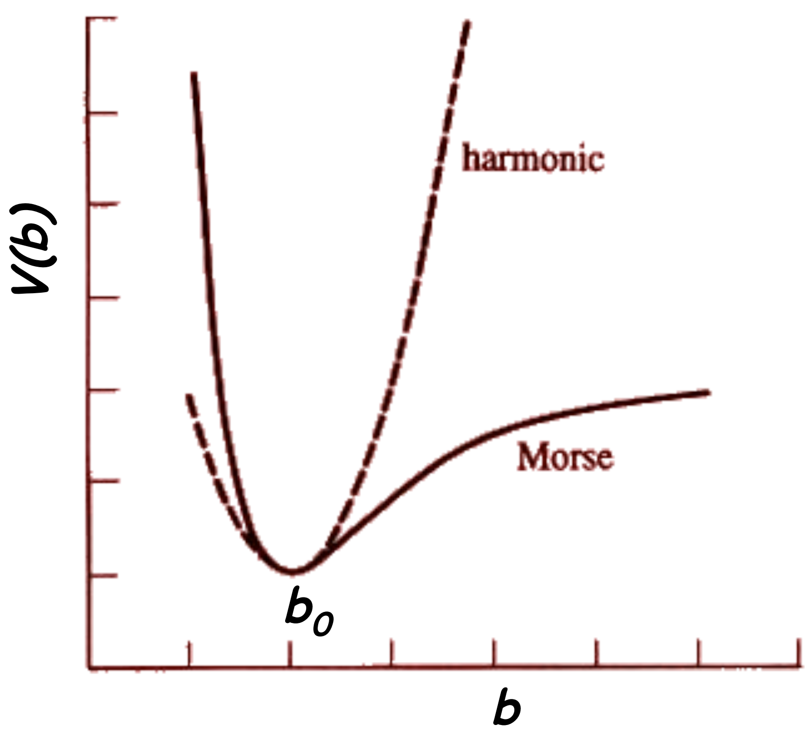
$$V(\vec{r}(t); D; K^b; b^0) = \sum_n D_n \left\{ 1 - e^{\left[-\left(\frac{K^b{}_n}{2} \right)^{1/2} (b_n(t) - b_n^0) \right]} \right\}^2$$

Allows dissociation
Computationally expensive

Harmonic potential

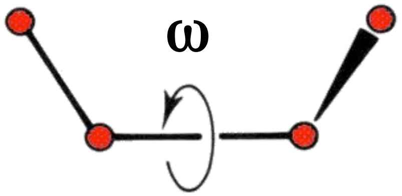
$$V(\vec{r}(t); K^b; b^0) = \sum_n \frac{1}{2} K_n^b (b_n(t) - b_0)^2$$

Most common form



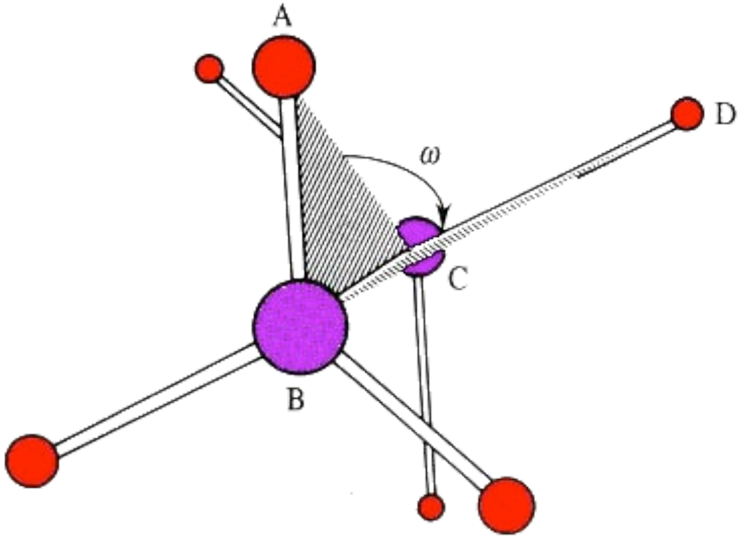
Atomic Interaction Functions

Dihedral angle or torsion interaction (4 body term



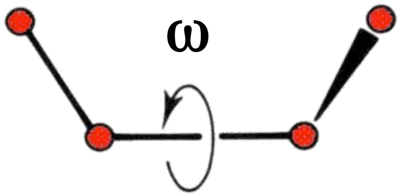
Trigonometric function:

- Describe rotation around bonds
- The torsion angle ω around the B-C bond is defined as the angle between the ABC and BCD planes
- $V_{torsion}$ should allow for multiple minima (rotameric states)



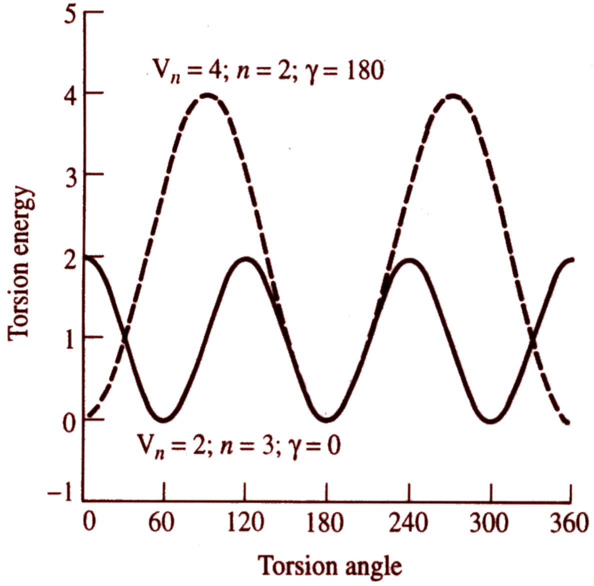
Atomic Interaction Functions

Dihedral angle or torsion interaction (4 body term)



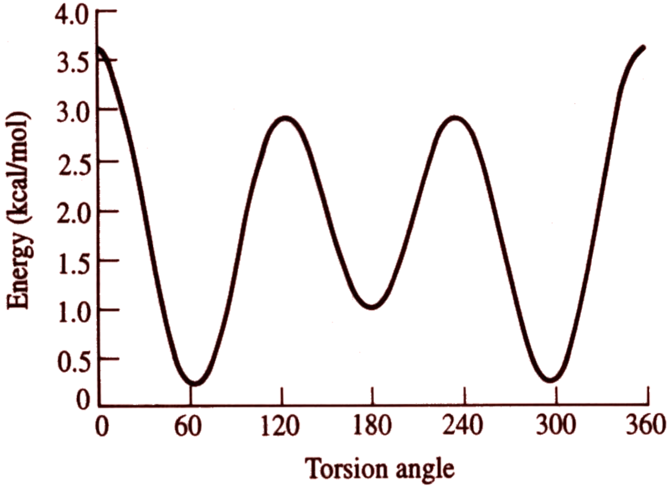
$$V_{dih}(\vec{r}(t);K^{dih};b^0) = \sum_n K_n^{dih}\left[1+\cos(n\omega_n(t)-\gamma\right]$$

$$\vec{F}_i(t) = K_n^{dih}\sin\left(n\omega_n(t)-\gamma\right)\frac{\partial}{\partial r_i(t)}\omega_n(t)$$



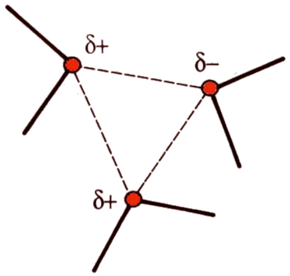
Single term: all minima equal

O-CH$_2$-CH$_2$-O



Two terms: minima no longer equal

Atomic Interaction Functions

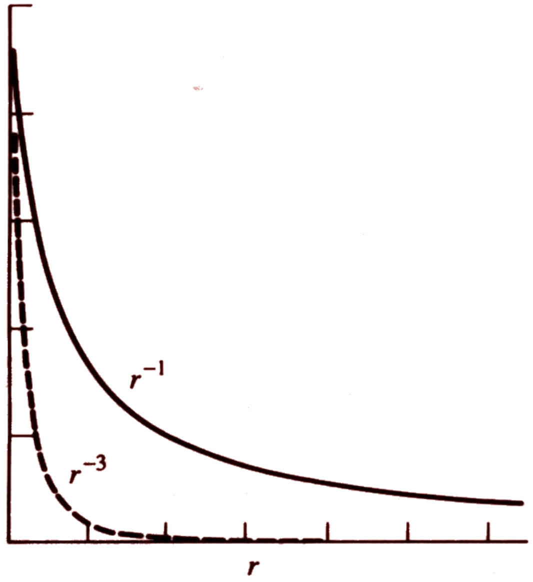


Electrostatic Interactions (treated as 2 body term; in principle N body)

- Electrostatic interactions calculated as the sum of inter-actions between pairs of point charges using Coulomb’s law
- Slow decay as function of distance between atoms (~ 1/r)
- Long-range contributions

$$V_{elec} = \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

q: partial charges
ε: dielectric constant



Non-Bonded Interactions

A. Lennard Jones + Coulomb

$$V(\vec{r}(t); B; A; q; \varepsilon_r) = \sum_{i<j} \left\{ \frac{B}{r_{ij}^{12}(t)} - \frac{A}{r_{ij}^{6}(t)} + \frac{1}{4\pi\varepsilon_o\varepsilon_r} \cdot \frac{q_i q_j}{r_{ij}(t)} \right\}$$

all pairs

Exclude 1st + 2nd neighbours
Modify 3rd neighbours

short range r_{ij} -> large V
bond and angle terms!

Forces

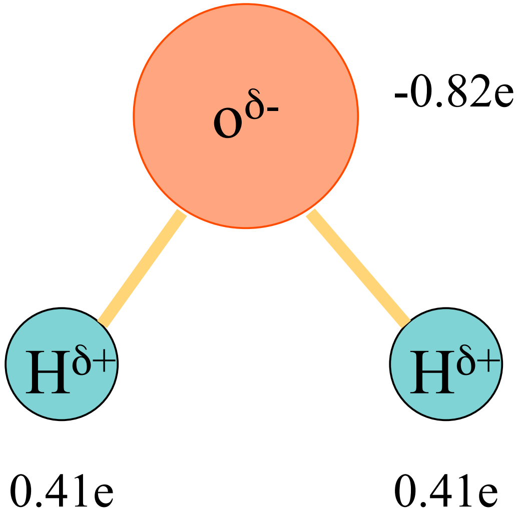
$$F = \left\{ \frac{12B}{r_{ij}^{12}(t)} - \frac{6A}{r_{ij}^{6}(t)} + \frac{1}{4\pi\varepsilon_o\varepsilon_r} \cdot \frac{q_i q_j}{r_{ij}(t)} \right\} \left[\frac{\vec{r}_{ij}}{r_{ij}^{2}} \right]$$

Parameterization

Coulomb Interactions

partial charge distributions

water



$$V = \frac{Z_i Z_j}{4\pi\varepsilon_0 r_{ij}}$$

Parameterization

Bonds	Length	Force Constant
	x-ray (small molecules) QM vibrational spectra (standard tables)	QM vibrational spectra empirically based on bond length weakly dependent geometry
Angles	Equilibrium Angle	Force Constant
	hybridization x-ray (small molecules) QM	QM vibrational spectra weakly dependent on bonds
Dihedral Angles	Multiplicity	Force Constant
	hybridization + substituents QM	fit to QM profiles (vacuum) population distributions (experimental) **strongly dependent on angles, vdw, charges, ...**

Parameterization (cont.)

van der Waals
self-terms
(Lennard-Jones)

Ar, Kr, Xe analytical expressions
(+ LJ phase diagram)

$C_6 \Rightarrow$ Slater-Kirkwood formula
C12 fit to density and pressure
(fixed temperature)

Fit to QM calculations
(not always reliable)

Cross-terms

Combination rules
fit to simple systems

vdw highly correlated with charge distribution

Parameterization (cont.)

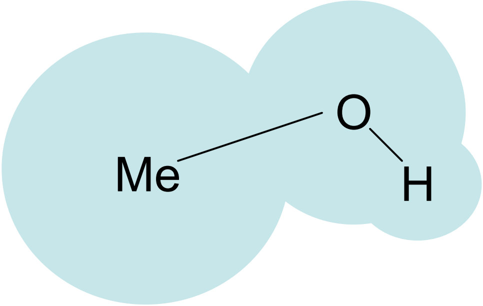
Partial atomic charges non-physical!

Partial Charges

- Fit to QM electron density calculations
- Estimate from electronegativities
- Fit to dipole moments
- Fit to hydrogen bonding energies
- Fit to properties of simple liquids

Dependent on dielectric, vdw's, treatment of long-range, all-atom/united atom, ...

Parameterization (cont.)



Methanol

Charges(e)	qH	qO	qMe
GR96	0.398	-0.574	0.176
H1	0.431	-0.728	0.297
B3	0.408	-0.674	0.266

Lennard-Jones $(10^{-3}$ kJmole/nm$^{12})^{1/2}$	$(C_{12}{}^{1}(O))^{1/2}$	$(C_{12}{}^{2}(O))^{1/2}$	$(C_{12}{}^{1}(Me))^{1/2}$
GR96	1.1250	1.227	4.5665
H1	1.5839	1.4683	5.7685
B3	1.5250	1.5250	4.4000

density/pressure/heat of vaporization similar (compensation of errors)

Compare to heat of vaparization, mixing enthalpy, diffusion, compressibility, dielectric constant, Debye relaxation times, solvation free energy, ...

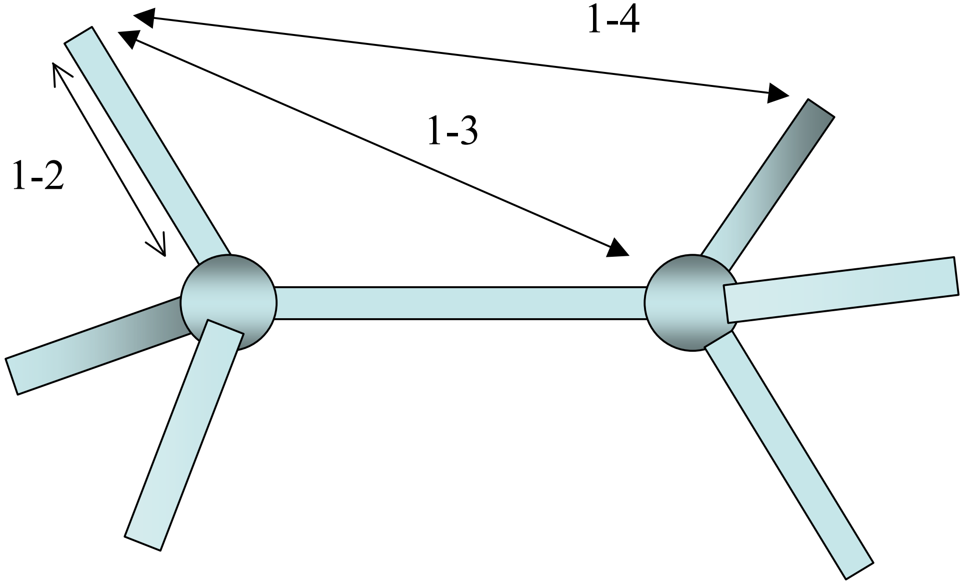
all properties cannot be reproduce simultaneously (polarization, Lennard Jones)

Relative Magnitude of Interactions

Interaction	**Dependence**	**Approximate magnitude (kJ/mole/nm)**
ion-ion	$\propto 1/r$	60
ion-dipole	$\propto 1/r^2$	~8
dipole-dipole	$\propto 1/r^3$	~2
dispersion	$\propto 1/r^6$	~.1

Other Considerations

Exclusions: Should we include all interactions?



1-2 excluded
1-3 excluded
1-4?

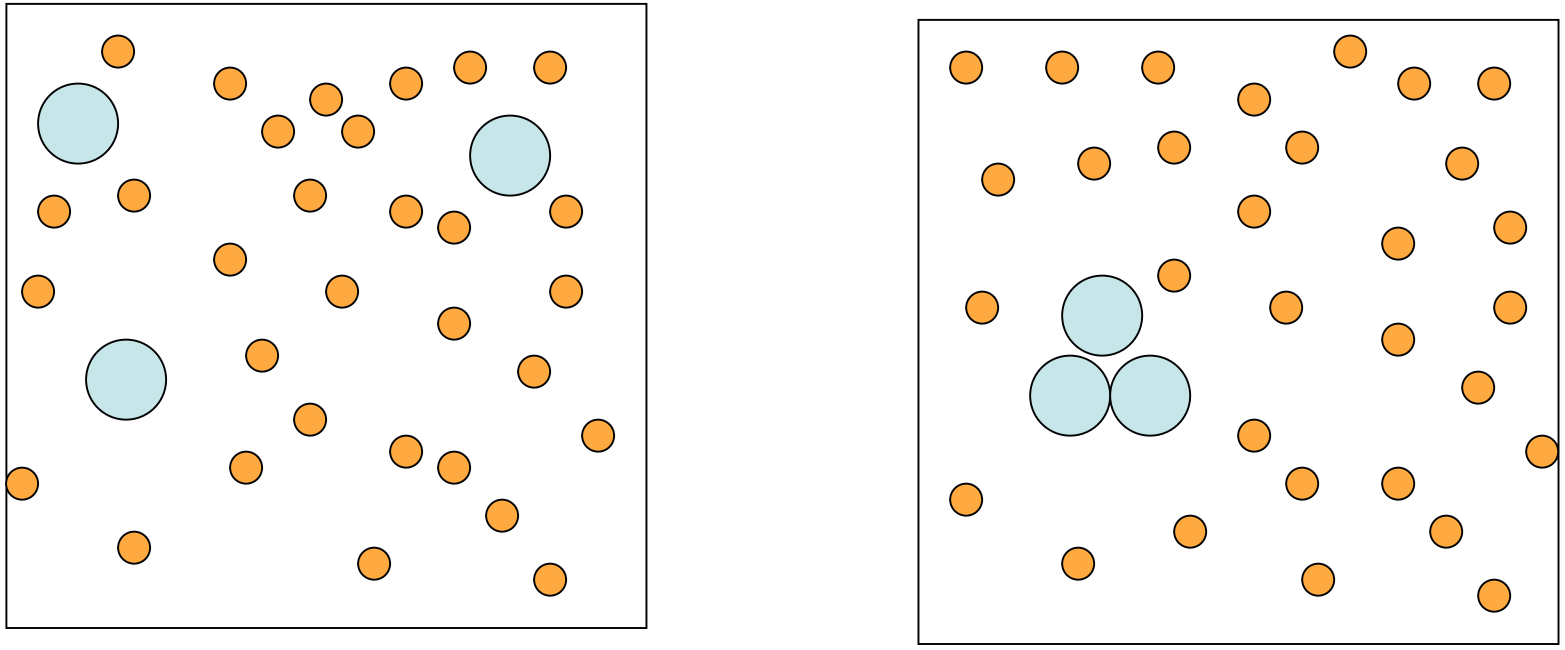
1-2 bonds
1-3 angles
1-4 dihedral (+ LJ and angle)

Effective interactions

Treatment varies between force fields

Hydrophobic Effect

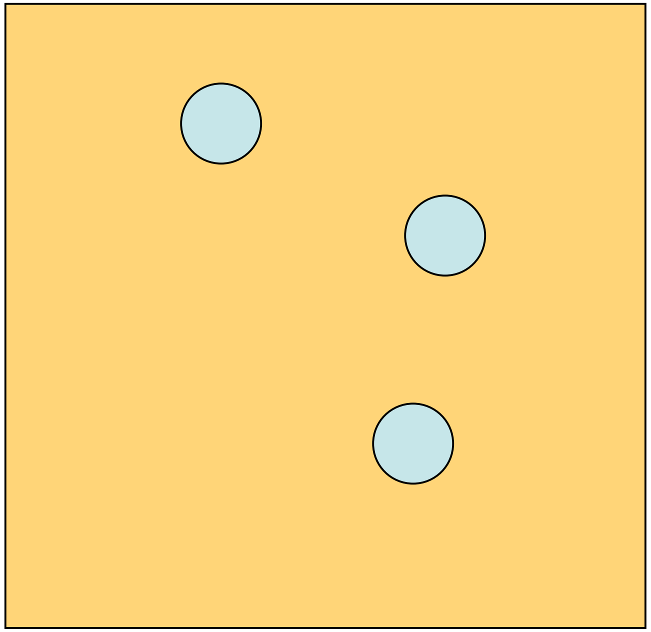
What keeps proteins folded?
Why does oil separate from water?
Why do people who do not know anyone at parties end up together?

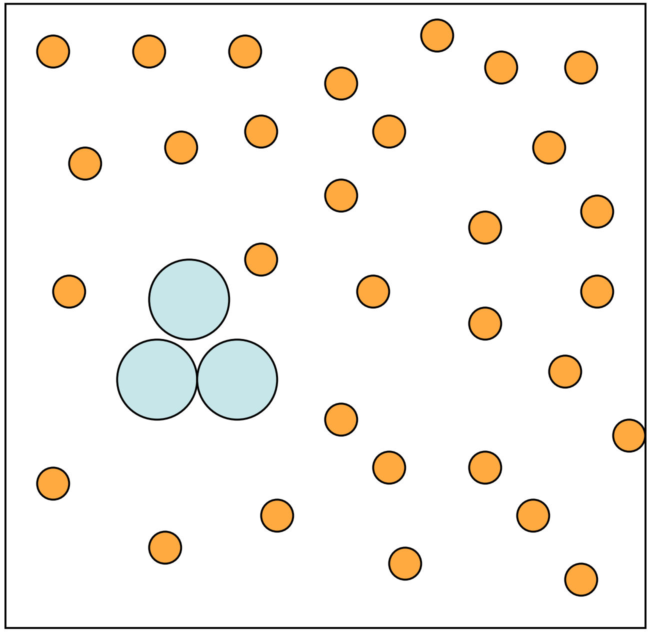


Particles driven together by favorable interaction within environment

not pairwise additive

Hydrophobic Effect


implicit solvent no-aggregation


Explicit solvent aggregation

NOT that particles do not like water. Rather the interaction of water with water stronger.

Entropy of solvent

Sometimes modeled as surface area term

Automatic in explicit solvent simulations

Empirical force fields

- **No “universal force field”**

- **Specific force fields are tailored to:**
  - *A given system*
  - *A given phase*
  - *A given property to be studied*
  - *A given computer budget*

- **Force field parameters are not physical constants. They are empirical parameters which are dependent (correlated):**
  - *On each other*
  - *On the functional form of the potential function*
  - *On the degrees of freedom*
  - *On the force field training set*

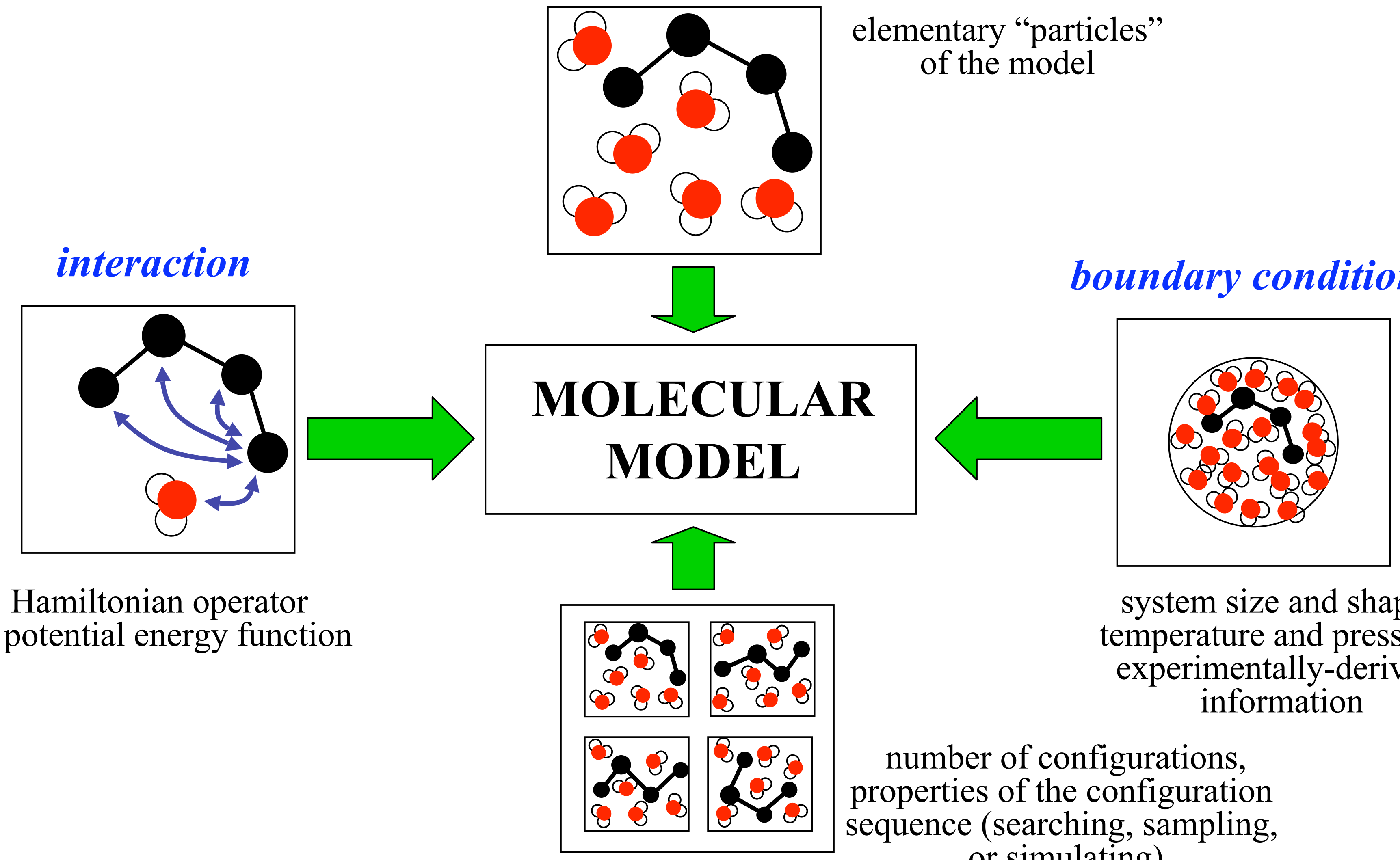
- **Force field parameters are generally not transferable from one force field to another**

- **The quality of a force field is limited by the crudest approximation made in its definition!**

A Common Force Fields

- Gromos96 **(*J. Comp. Chem.* 19, 535 (1998))**
  - http://www.igc.ethz.ch/gromos-docs/index.html
- AMBER **(*J. Am. Chem. Soc.*** 117**, 5179 (1995))**
  - http://www.amber.ucsf.edu/amber/amber.html
- OPLS **(*J. Phys. Chem B*,** 105**, 6474 (2001))**
  - http://zarbi.chem.yale.edu
- CHARMM (***Encyclopedia of Comp. Chem.***, 1, **271. (John Wiley & Sons 1998))**
  - http://www.scripps.edu/brooks

FOUR BASIC CHOICES DEFINING A MOLECULAR MODEL



degrees of freedom

elementary “particles”
of the model

interaction

Hamiltonian operator
or potential energy function

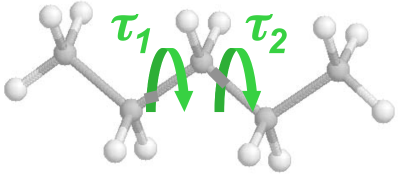
MOLECULAR
MODEL

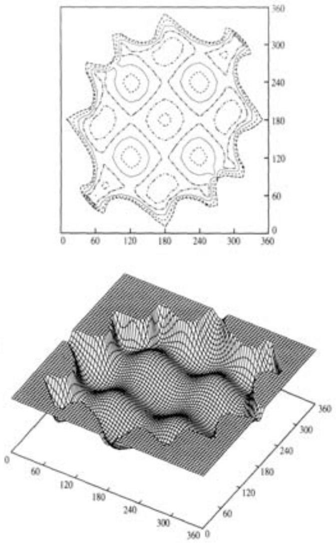
boundary conditions

system size and shape,
temperature and pressure,
experimentally-derived
information

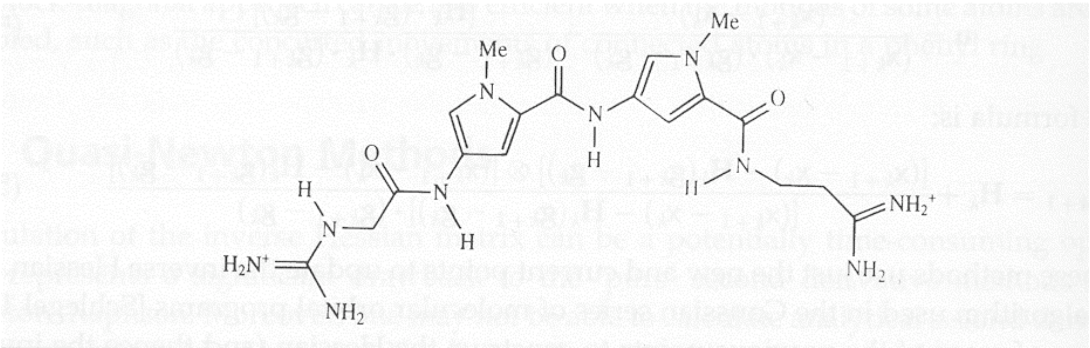
generation of configurations

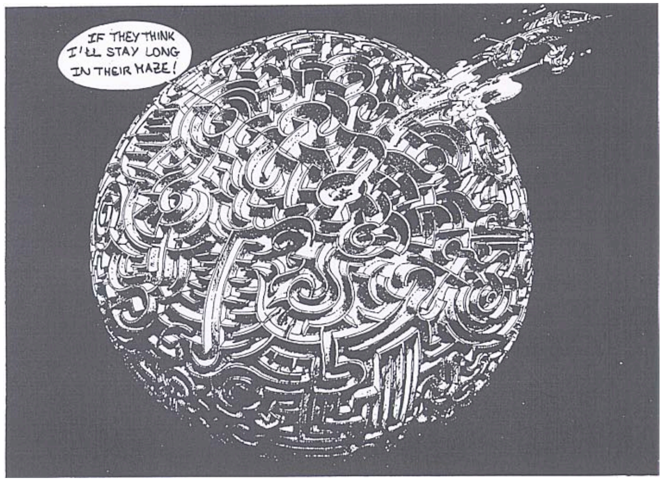
number of configurations,
properties of the configuration
sequence (searching, sampling,
or simulating)





Searching Configurational Space





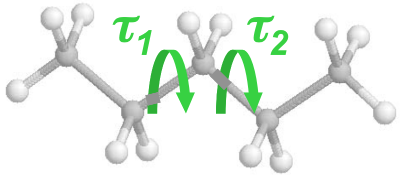
POTENTIAL ENERGY SURFACES

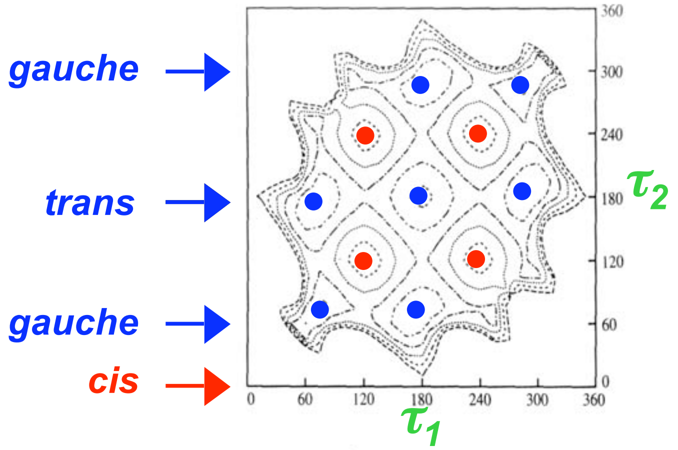
Potential energy surface (PES)

force field
or
quantum-chemical method
(within Born-Oppenheimer)

} potential energy (hyper)surface defined by $V(\boldsymbol{r}) = V(\{\boldsymbol{r}_i\})$ with $3N$ (Cartesian) or $3N$-6 (internal) dimensions

Example: pentane

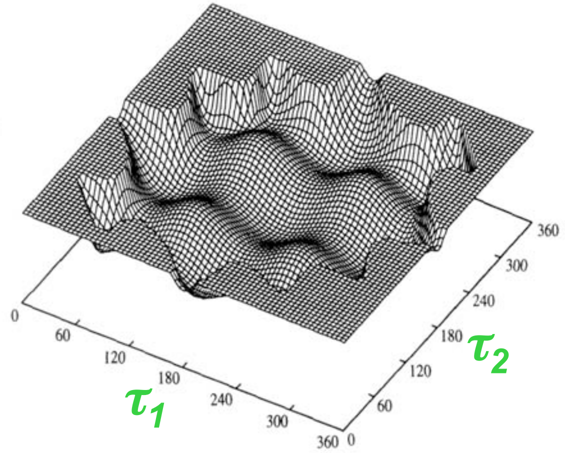




rigid bonds and bond-angles
optimized methyl orientation

⇒ 2 degrees of freedom

⇒ 9 minima and 9 maxima



How to find the minimum energy conformation?

POTENTIAL ENERGY SURFACES (2)

Biomolecular Systems

- many dimensions

⇒ *no hope to enumerate all configurations*

e.g. ***100 residue protein***,
[~15 atoms per residue]
→ ***4500 degrees of freedom***

- many energy minima (the PES is "*frustrated*")

⇒ *no hope to enumerate all minimas*

e.g. ***alkane with n carbons***
[~ 3 minima per torsion]
→ ***about*** $\boldsymbol{3^n}$ ***minima***
(n=10: 59'049, n=20: 3'486'784'401)

Generating configurations

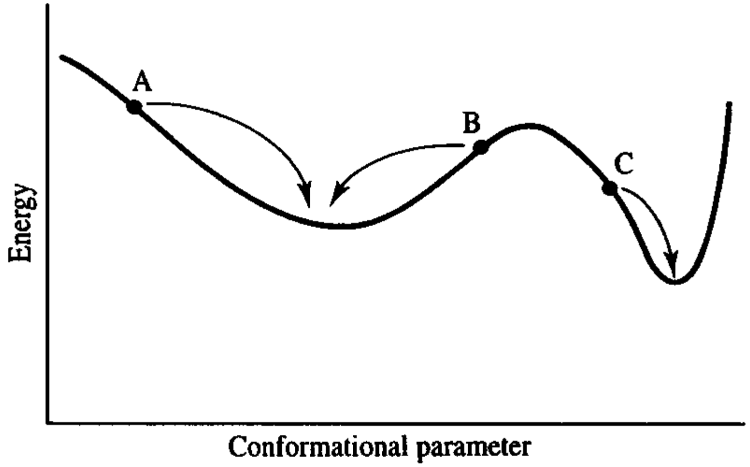
- Methods must preferentially generate ***relevant configurations***
  (*e.g.* low energy, Boltzmann-weighted)

- Need good ***initial configuration***
  (*e.g.* from X-ray or NMR experiments)

Method

- If ***Boltzmann-weighted ensemble*** of configurations⇒ ***thermodynamic properties***
- if configurations generated using ***equations of motion*** ⇒ ***dynamic properties***

Energy Minimization

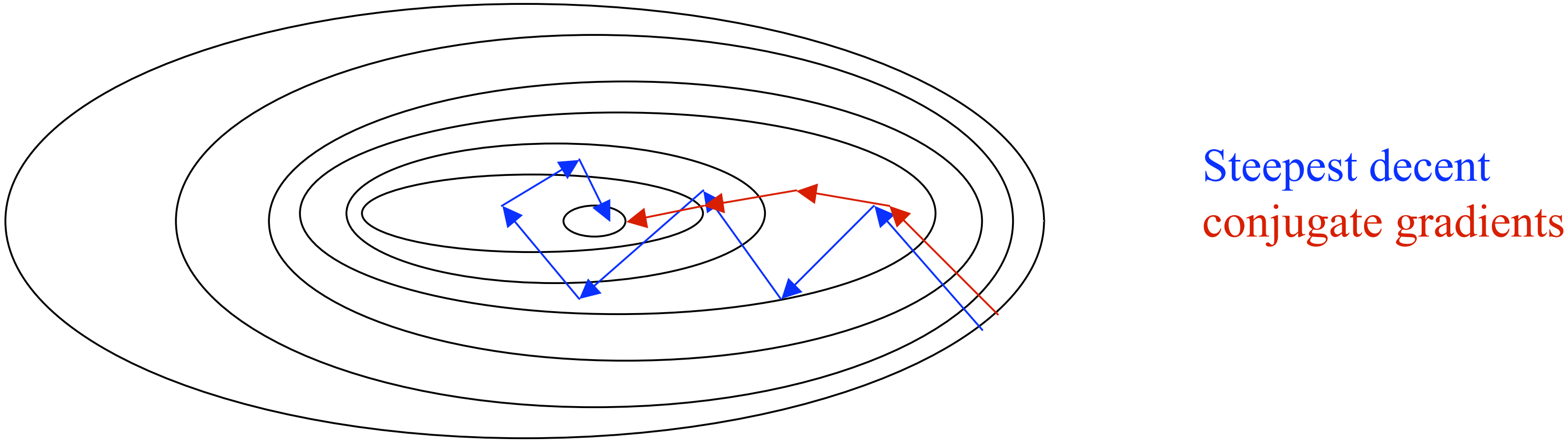


- EM methods:
  - – Iterative approach to complex problems
  - – Succession of downhill moves
  - – Find the minimum closest to an initial configuration r_o
  - – Poor search method in complex systems:
    - *Applied to a large set of initial low-energy configurations r_o*
    - *Combined with uphill moves*
  - – The global minimum seldom located!

Energy minimization

- Non-derivative methods
  - Only need the energy V of the system
  - e.g. simplex method
- Derivative methods (need gradient $g_i = dV/dr_i$)
  - Analytical or numerical (finite difference, expensive!) derivatives
  - e.g. steepest descents, conjugate gradients, Newton-Raphson, …

Energy Minimization



i) Steepest decent (SD): step along gradient
step size: V becomes lower ⇒ increase step size
V becomes higher ⇒ decrease step size

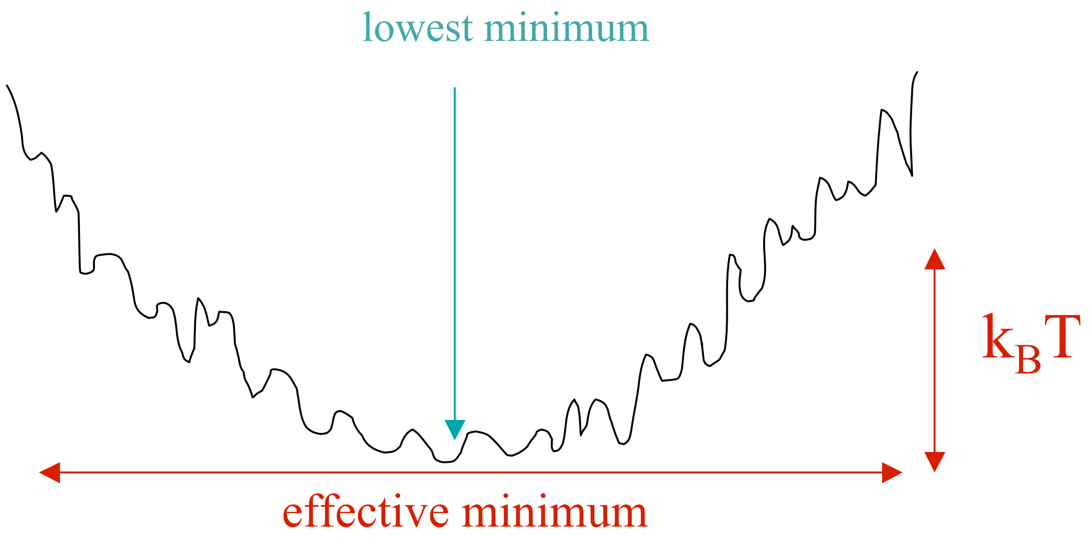
ii) Conjugate gradients (CG): step along linear combination of current and previous gradients
step size: from local quadratic approximation

Application of Energy Minimization

Relieve strain or very high energies from the system with steepest decent or conjugate gradients (very useful!)

Use of 2nd order or quasi-Newtonian methods?

In general overkill: At T~300K not interested in the exact location of minimum



lowest minimum

$k_B T$

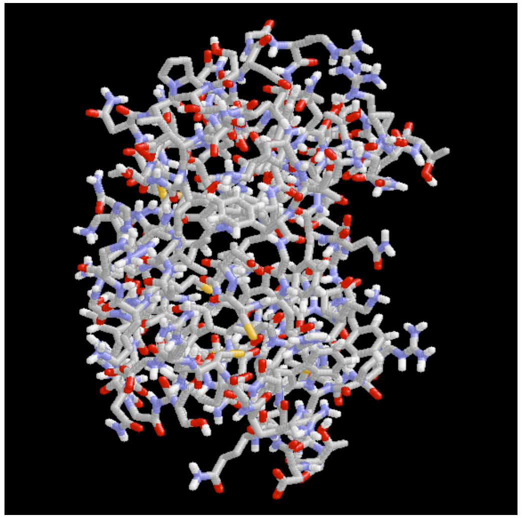
effective minimum

EM to search conformational space?

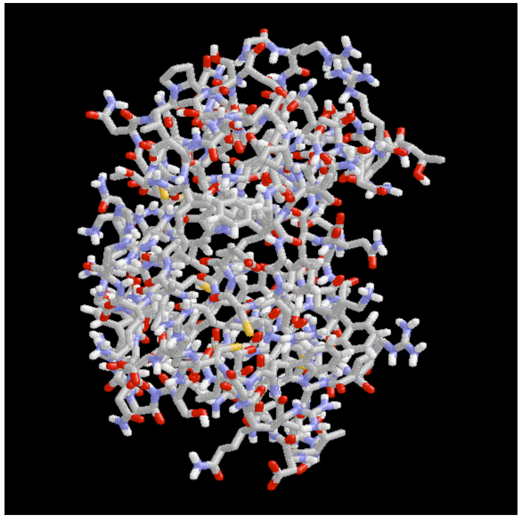
No: MD, SD much more efficient

ENERGY MINIMA & FRUSTRATED SYSTEMS

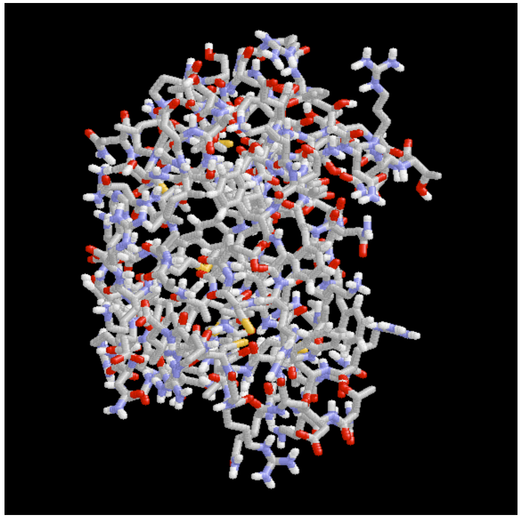
e.g. hen-egg-white lysozyme in vacuum



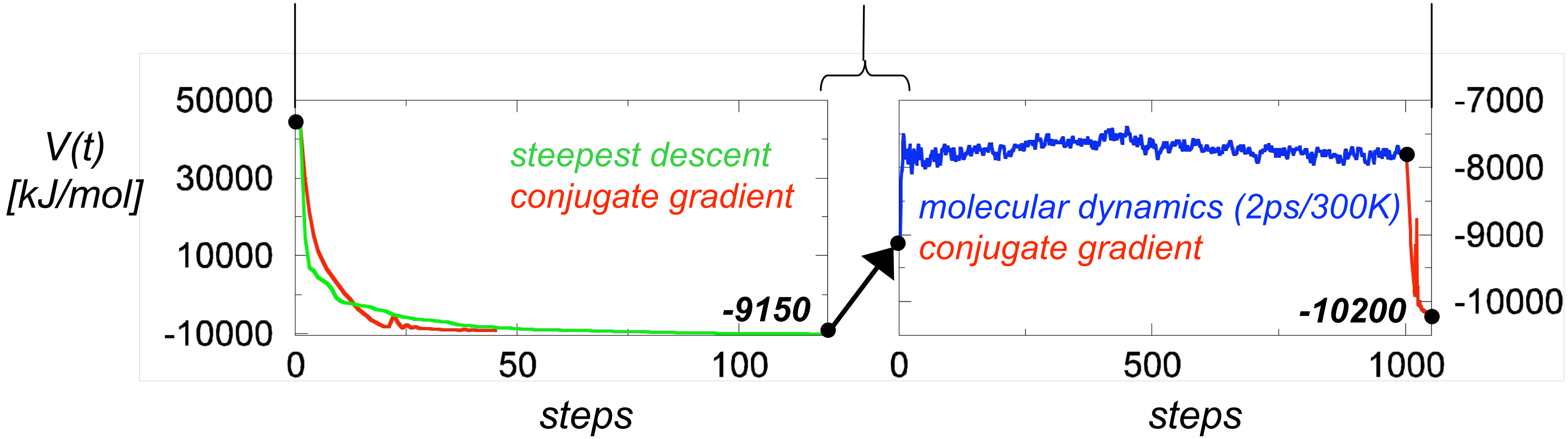
X-ray structure



after EM

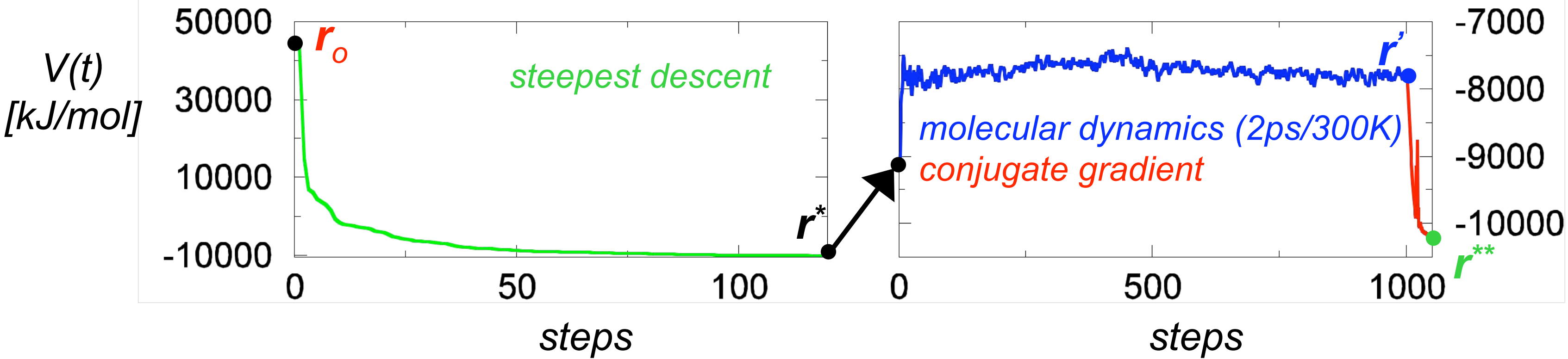


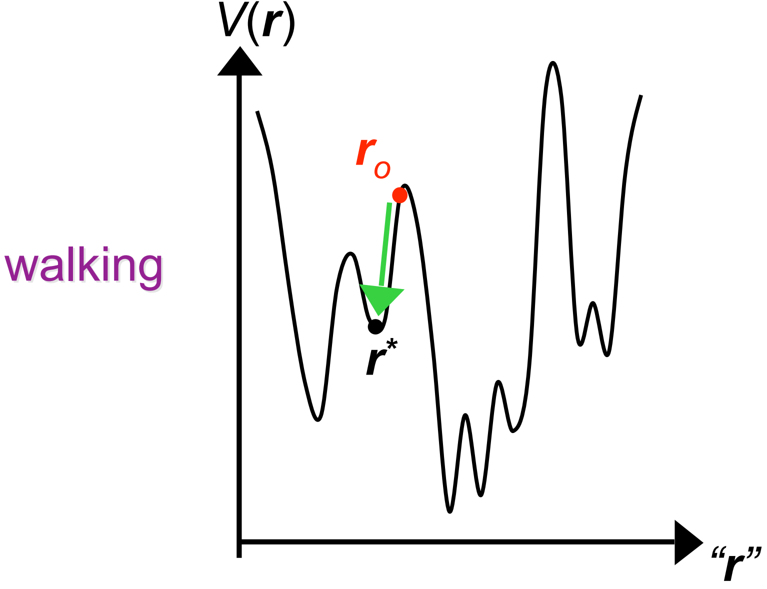
after MD and EM

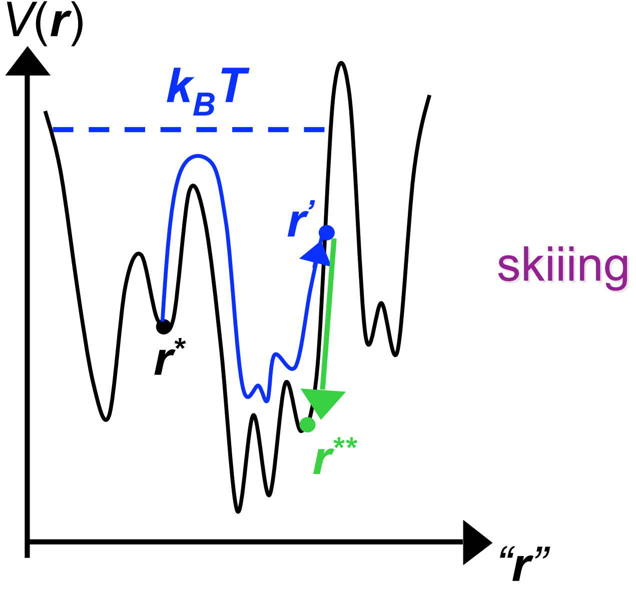


ENERGY MINIMA & FRUSTRATED SYSTEMS (2)

Explanation







⇒ energy minimization is a *poor search method* (lacks uphill moves)

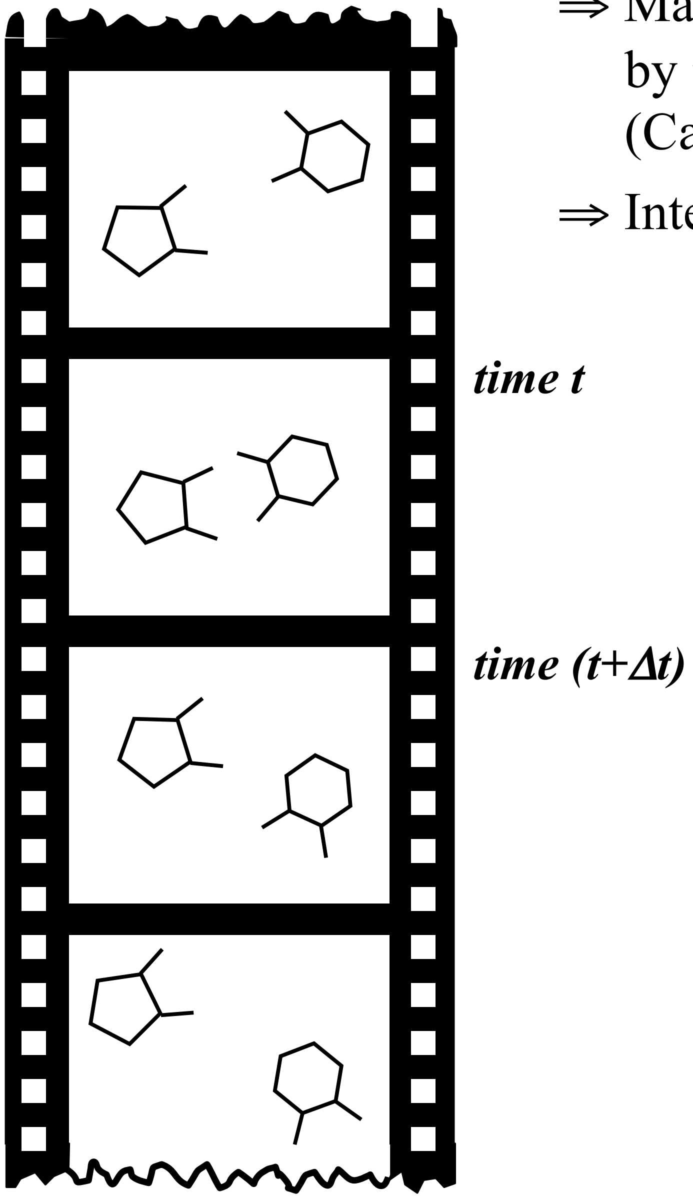
⇒ used to *relax strain and/or* generate input for dynamics simulations

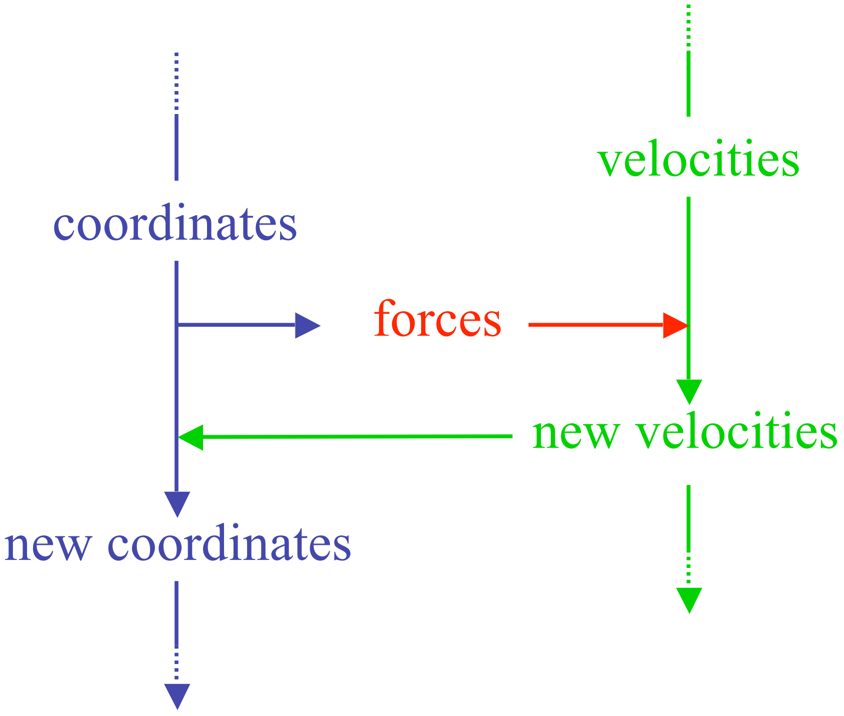
Molecular Dynamics

MOLECULAR DYNAMICS (PRINCIPLE)

⇒ Making a „movie“ of a molecular system by integrating the classical equations of motion (Cartesian coordinate system → Newton formulation)

⇒ Integration timestep $\Delta t \cong 10^{-15}$ seconds !





velocities

coordinates

forces

new velocities

new coordinates

Newton's Equations of Motion

force = mass x acceleration

or

acceleration = force/mass

$$\frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i}$$

Force = -slope of the potential

$$F_i = \frac{\partial}{\partial r_i} V(r_1, r_2, \ldots, r_N)$$

Note:

1. Valid in Cartesian coordinates systems only
   (other coordinate systems Lagrangian or Hamiltonian formulations).
2. Conservation of energy (in conservative systems).
3. Deterministic
4. Time reversible.

Some Basic Definitions (Classical Mechanics)

System of N particles $\vec{r}_1, \vec{r}_2, \cdots, \vec{r}_N$

Cartesian coordinates $\{\vec{r}_1, \vec{r}_2, \cdots, \vec{r}_N\} \equiv \vec{r}^{\,N} \equiv \vec{r}$

Interaction function or energy $V(\vec{r}_1, \vec{r}_2, \cdots, \vec{r}_N)$

Force on particle, *i* $\vec{F}_i = \vec{\nabla}_i V \equiv \frac{\partial V}{\partial \vec{r}_i} \equiv \left(\frac{\partial V}{\partial x_i}, \frac{\partial V}{\partial y_i}, \frac{\partial V}{\partial z_i} \right)$

Newton's equations of motion $\vec{F}_i = m_i \ddot{\vec{r}}_i \Leftrightarrow m_i \dot{\vec{v}}_i = \vec{F}_i ; \dot{\vec{r}}_i = \vec{v}_i$

Kinetic energy

$$K = \sum_{i=1}^{N} \frac{1}{2} m_i \dot{\vec{r}}^{\,2}{}_i = \sum_{i=1}^{N} \frac{\vec{p}^{\,2}}{2m_i}$$

$$= \sum_{i=1}^{N} \frac{1}{2} m_i \left[\dot{x}_i + \dot{y}_i + \dot{z}_i \right]$$

mass $m_1, m_2, \cdots, m_N$

Velocity $\vec{v}_i \equiv \frac{d}{dt} \vec{r}_i \equiv \dot{\vec{r}}_i$

Acceleration $\vec{a}_i \equiv \frac{d}{dt} \vec{v}_i \equiv \ddot{\vec{r}}_i$

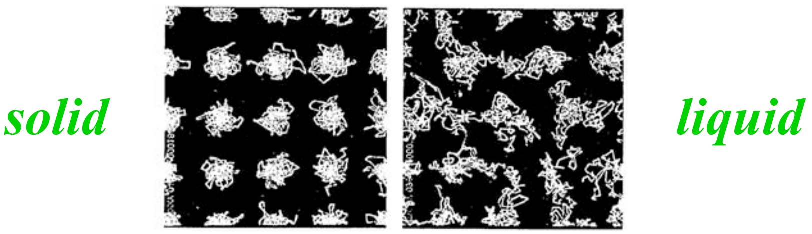
Momentum $\vec{p}_i = m_i \vec{v}_i$

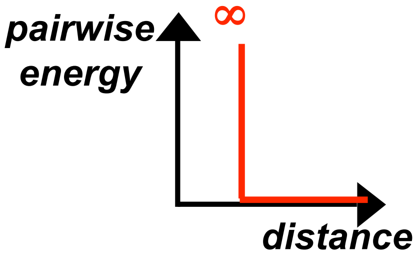
Potential energy $V(\vec{r}_1, \vec{r}_2, \cdots, \vec{r}_N)$

MOLECULAR DYNAMICS (INTEGRATOR)

A primitive integrator

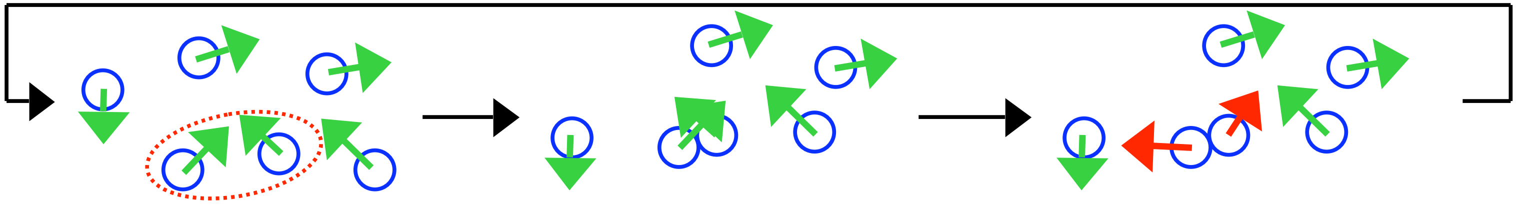
⇒ Alder & Wainwright 1957 – first MD study of hard spheres in the condensed phase





⇒ Integrator

(sampling is not uniform in time)



locate next colliding pair and determine collision time

compute coordinates at collision time

update velocities of colliding particles (elastic choc)

Algorithms for Molecular Dynamics

Equation

(What do we have to solve)

$\ddot{x}(t) = f(x)$ with $\dot{x}(t) = v(x)$, $\dot{v}(t) = f(x)$

integrate f(x) in small steps with $t_n = n\Delta t$

1. Open Forms:

A. Euler Methods

(most basic)

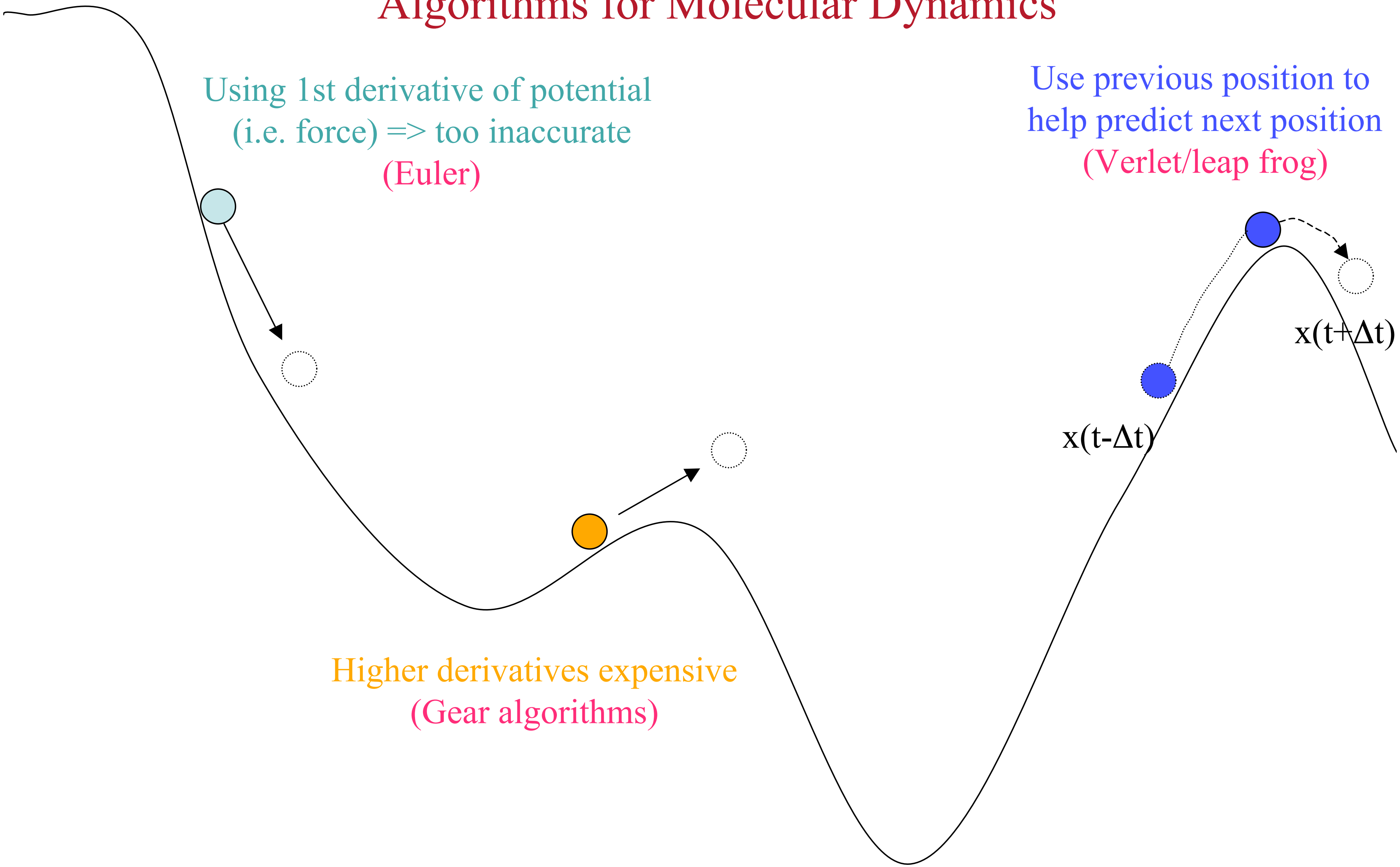
$$x_{n+1} = x_n + v_n \Delta t + \frac{1}{2} m^{-1} F(x_n) \Delta t^2$$

$$v_{n+1} = v_n + m^{-1} F(x_n) \Delta t$$

no derivative of force

no correction for the curvature of the potential energy surface

Algorithms for Molecular Dynamics



Using 1st derivative of potential
(i.e. force) => too inaccurate
(Euler)

Higher derivatives expensive
(Gear algorithms)

Use previous position to
help predict next position
(Verlet/leap frog)

The Verlet Algorithm for Integrating Newtons Equations of Motion:

Newton: $\ddot{\vec{r}}_i = m_i^{-1}\vec{F}_i \quad \Rightarrow$

$$\frac{d\vec{r}}{dt} = \vec{v} \qquad \textit{velocity}$$

$$\frac{d\vec{v}}{dt} = m^{-1}\vec{F}(\vec{r}(t))$$

force

Taylor expansion for a small time step, Δt.

Taylor expansion: $f(x+h) = f(x) + hf'(x) + \frac{h^2}{2!}f''(x) + \frac{h^3}{3!}f'''(x) + \ldots$

Expand $\vec{r}(t_n + \Delta t)$ and $\vec{r}(t_n - \Delta t)$

$$\vec{r}(t_n + \Delta t) = \vec{r}(t_n) + \left.\frac{d}{dt}\vec{r}\right|_{t_n}\Delta t + \frac{1}{2!}\left.\frac{d^2}{dt^2}\vec{r}\right|_{t_n}\Delta t^2 + \frac{1}{3!}\left.\frac{d^3}{dt^3}\vec{r}\right|_{t_n}\Delta t^3 + O(\Delta t^4)$$

$$\vec{r}(t_n - \Delta t) = \vec{r}(t_n) - \left.\frac{d}{dt}\vec{r}\right|_{t_n}\Delta t + \frac{1}{2!}\left.\frac{d^2}{dt^2}\vec{r}\right|_{t_n}\Delta t^2 - \frac{1}{3!}\left.\frac{d^3}{dt^3}\vec{r}\right|_{t_n}\Delta t^3 + O(\Delta t^4)$$

Sum $\vec{r}(t_n + \Delta t)$ and $\vec{r}(t_n - \Delta t)$ to express the new position $\vec{r}(t_n + \Delta t)$ in terms of the previous position $\vec{r}(t_n - \Delta t)$ and the current position $\vec{r}(t_n)$

$$\vec{r}(t_n + \Delta t) + \vec{r}(t_n - \Delta t) = 2\vec{r}(t_n) + \left.\frac{d^2}{dt^2}\vec{r}\right|_{t_n}\Delta t^2 + O(\Delta t^4)$$

$$\vec{r}(t_n + \Delta t) = 2\vec{r}(t_n) - \vec{r}(t_n - \Delta t) + m^{-1}\vec{F}(\vec{r}(t_n))\Delta t^2 + O(\Delta t^4)$$

The Verlet Algorithm for Integrating Newtons Equations of Motion (cont.)

Take difference between $\vec{r}(t_n + \Delta t)$ and $\vec{r}(t_n - \Delta t)$

$$\vec{r}(t_n + \Delta t) - \vec{r}(t_n - \Delta t) = 2\frac{d}{dt}(\vec{r})\bigg|_{t_n} \Delta t + \frac{1}{3!}\frac{d^3}{dt^3}\vec{r}\bigg|_{t_n} \Delta t^3 + O(\Delta t^5)$$

Use difference to calculate velocities

$$\vec{v}_n(t_n) = \frac{\vec{r}(t_n + \Delta t) - \vec{r}(t_n - \Delta t)}{2\Delta t} - \frac{\vec{F}(t_{n+1}) - \vec{F}(t_{n-1})}{12}\frac{\Delta t^2}{m}$$

$$\frac{d}{dt}\vec{F}\bigg|_{t_n} \approx \frac{\vec{F}(t_{n+1}) - \vec{F}(t_{n-1})}{2\Delta t}$$

Note: Using the Verlet algorithm the trajectory is independent of the velocities: Coupling to a T-bath is NOT possible.

$$\frac{3}{2}k_B T = \frac{1}{2}m_i {v^2}_i$$

The Leap-Frog Algorithm for Integrating Newton's Equations of Motion

Newton: $\ddot{\vec{r}}_i = m_i^{-1}\vec{F}_i$ → 1-dimensional

$$\begin{cases} \dfrac{dx}{dt} = v & \text{velocity} \\ \dfrac{dv}{dt} = m^{-1}F(x) & \\ \quad\text{force} & \end{cases}$$

$$v\left(t_n + \frac{\Delta t}{2}\right) = v(t_n) + \left.\frac{dv}{dt}\right|_{t_n}\frac{\Delta t}{2} + \frac{1}{2!}\left.\frac{d^2v}{dt^2}\right|_{t_n}\left(\frac{\Delta t}{2}\right)^2 + O(\Delta t^3)$$

$$v\left(t_n - \frac{\Delta t}{2}\right) = v(t_n) - \left.\frac{dv}{dt}\right|_{t_n}\frac{\Delta t}{2} + \frac{1}{2!}\left.\frac{d^2v}{dt^2}\right|_{t_n}\left(\frac{\Delta t}{2}\right)^2 - O(\Delta t^3)$$

Taylor expansion for a small time step, Δt.

Velocity:

$$v\left(t_n + \frac{\Delta t}{2}\right) - v\left(t_n - \frac{\Delta t}{2}\right) = \left.\frac{dv}{dt}\right|_{t_n}\Delta t + O(\Delta t^3)$$

$$v\left(t_n + \frac{\Delta t}{2}\right) = v\left(t_n - \frac{\Delta t}{2}\right) + m^{-1}F(x(t_n))\Delta t + O(\Delta t^3)$$

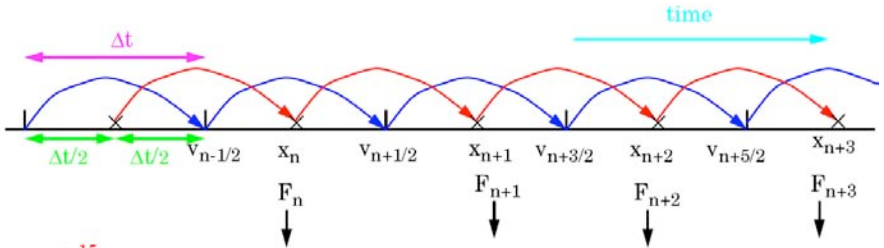
Take difference

Analogous for x:

$$x\left(t_n + \frac{\Delta t}{2}\right) - x\left(t_n - \frac{\Delta t}{2}\right) = \left.\frac{dx}{dt}\right|_{t_n}\Delta t + O(\Delta t^3)$$

$$x\left(t_n + \frac{\Delta t}{2}\right) = x\left(t_n - \frac{\Delta t}{2}\right) + v(t_n)\Delta t + O(\Delta t^3)$$

Shift time origin: $x(t_n + \Delta t) = x(t_n) + v\left(t_n + \frac{\Delta t}{2}\right)\Delta t + O(\Delta t^3)$



$\Delta t \approx 10^{-15}$ sec
0.002 ps

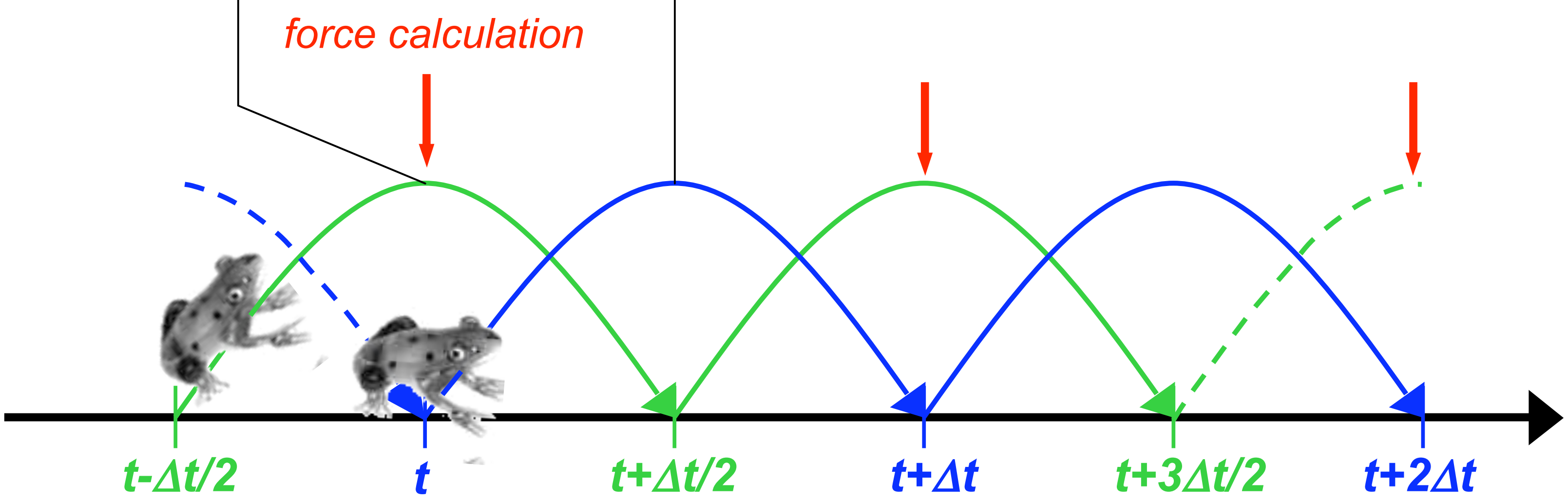
Trajectory of the particles: write to disk

MOLECULAR DYNAMICS (LEAP-FROG)

The leap-frog integrator in practice

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

$$\mathbf{r}(t+\Delta t)=\mathbf{r}(t)+\mathbf{v}(t+\frac{1}{2}\Delta t)\Delta t$$



Coordinates and velocities are not available simultaneously

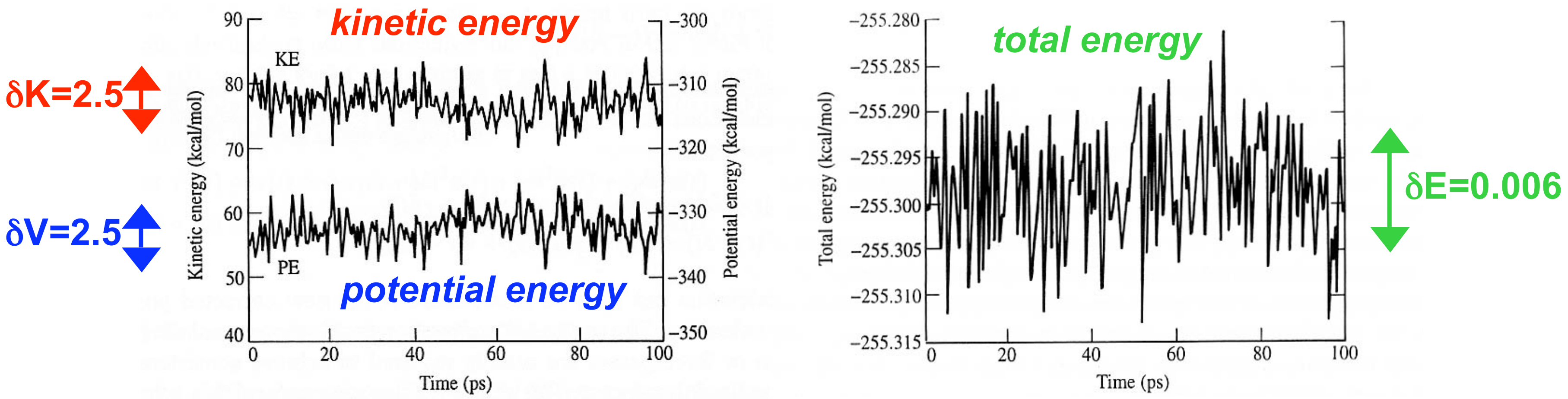
⇒ to get the kinetic energy at time t, the corresponding velocities must be back-calculated

$$\mathbf{v}(t)=\frac{1}{2}\left[\mathbf{v}(t-\frac{1}{2}\Delta t)+\mathbf{v}(t+\frac{1}{2}\Delta t)\right]+\frac{\Delta t}{16}\left[\mathbf{a}(t-\Delta t)-\mathbf{a}(t+\Delta t)\right]+O[(\Delta t)^4]$$

MOLECULAR DYNAMICS (ENERGY CONSERVATION)

Energy conservation

$\Rightarrow$ liquid argon (256 atoms), ρ=1.396 g/cm^3, T_{init}=100K, Velocity-Verlet, Δt=10 fs



$\Rightarrow$ the kinetic and potential energies fluctuate significantly

$\Rightarrow$ the total energy is essentially conserved

- in the absence of non-conservative forces (not always true in practice !), it should be exactly conserved in the limit $\Delta t \rightarrow 0$
- at finite Δt energy conservation is limited by the integration accuracy

  $\rightarrow$ intrinsic accuracy of the integrator (neglect of terms of $O[(\Delta t)^n]$)

  $\rightarrow$ timestep size *e.g.* δE = 0.006 (Δt=10fs), δE = 0.002 (Δt=5fs), and δE = 0.040 (Δt=25fs)

Integration Time Step

Integration Time Step

The time step determines how much time can be simulated. The smaller the time step the more expensive the calculation

Kinetic energy = 1/2 mv^2

From equipartition kinetic energy = $1/2k_BT$ (per degree for freedom)

mass of atom ↓ velocity of atom ↑ need smaller timestep

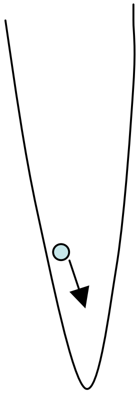
temperature ↑ velocity of atom ↑ need smaller timestep

Time step determined by lightest atom (highest frequency motion)

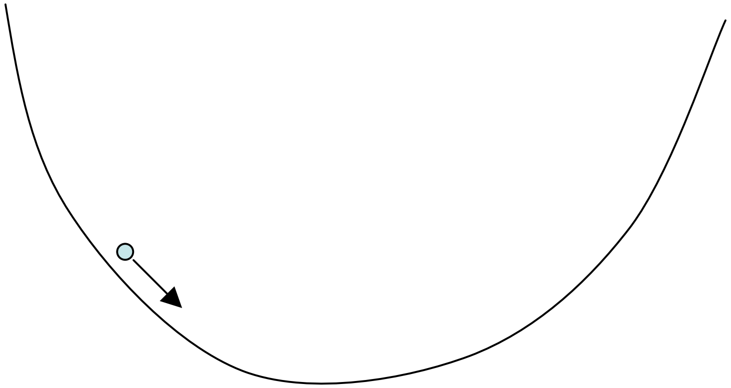
Integration Time Step (cont.)

The maximum time step determined by the curvature of the potential

(frequency of the motion)



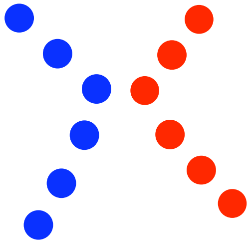
Steep potential small time step



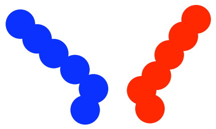
Shallow potential large time step

MOLECULAR DYNAMICS (TIMESTEP)

Choice of a timestep

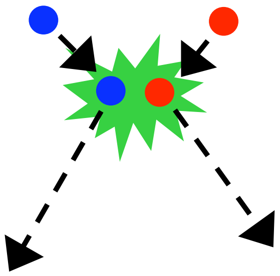


reasonable



too short

→ *poor sampling*



too large

→ *poor energy conservation*

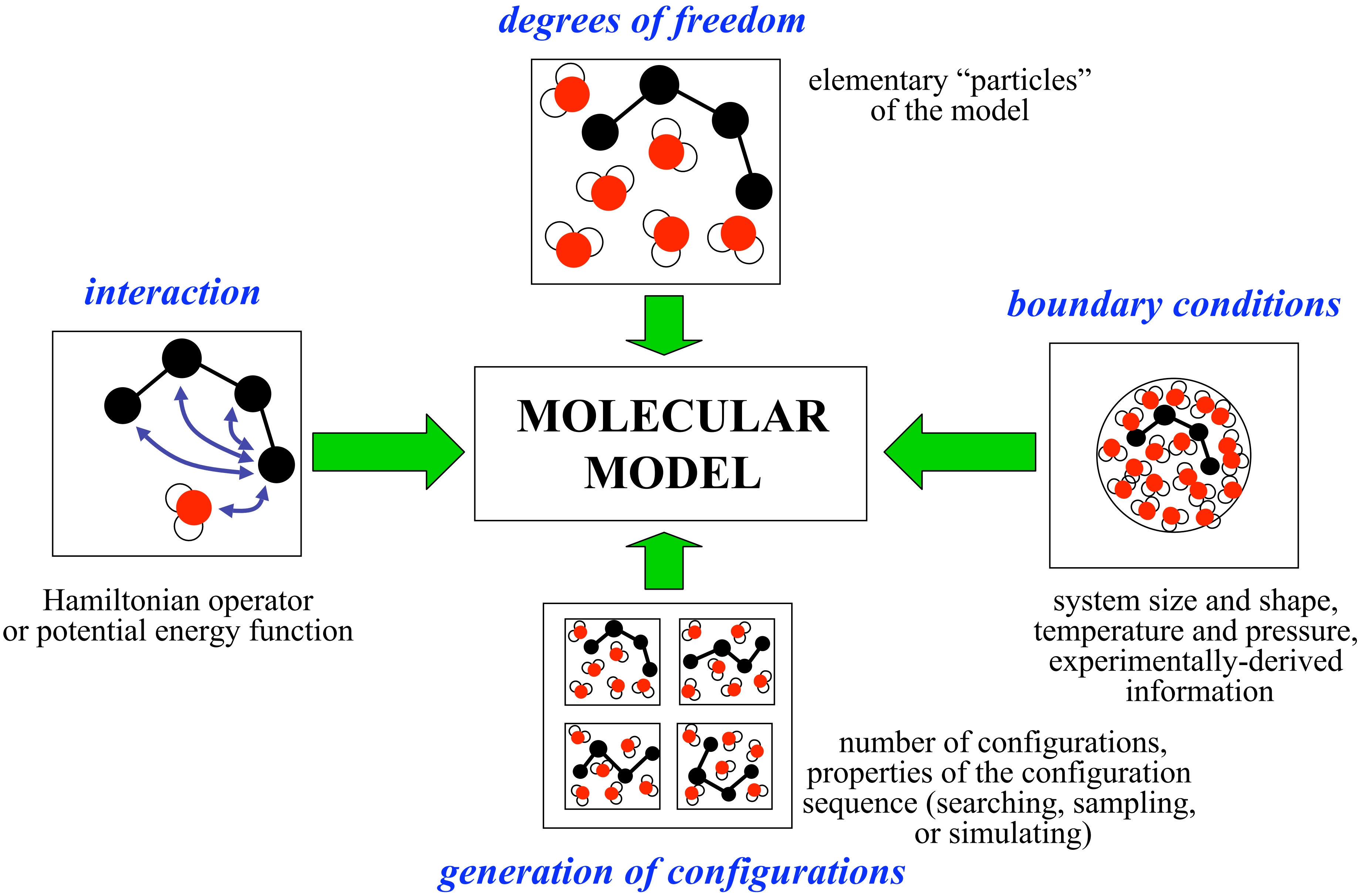
→ *program overflow or failure*

⇒ ***Rule of thumb:*** $\Delta t = \dfrac{\tau}{10}$

where τ is the period of the fastest motion in the system

System	***Motions***	***Timescale***	***Timestep***
Atomic liquid (*e.g.* argon)	translation (vibration at contact)	≤ 1000 fs (from LJ curve)	10 fs
Molecular liquid (rigid molecules)	*idem* + rotation	≤ 500 cm^{-1}	5 fs
Flexible molecules (rigid bonds)	idem + torsion + bond-angle vibration	≤ 2000 cm^{-1}	2 fs
Flexible molecules	*idem* + bond-stretching vibration	≤ 3000 cm^{-1} (C-H)	0.5-1 fs

FOUR BASIC CHOICES DEFINING A MOLECULAR MODEL



degrees of freedom

elementary "particles"
of the model

interaction

Hamiltonian operator
or potential energy function

MOLECULAR
MODEL

boundary conditions

system size and shape,
temperature and pressure,
experimentally-derived
information

number of configurations,
properties of the configuration
sequence (searching, sampling,
or simulating)

generation of configurations

Treatment of Boundaries

Q: Why worry about boundaries?
A: Number of atoms << $N_{avagadro} = 6.022 \times 10^{23}$

1. No wall:

Distortive effects!!

a) Surface tension → reduction in surface area.
Molecules become more spherical.
DNA, Insulin deform
myoglobin, cytochrome deform less (more globular).

Partial remedy: Use solvent to make system more spherical *(pressure effects)*.

b) Dielectric permittivity of vacuum = 1
Charge-charge interactions in vacuum >> polar solvents
all solvents $\varepsilon > 1$
water $\varepsilon = 78$

partial remedy: Reduce partial charges
Change ε

Treatment of Boundaries

2. Extended Wall Region:

Restrict motion of atoms in a layer adjacent to vacuum to inhibit distortion

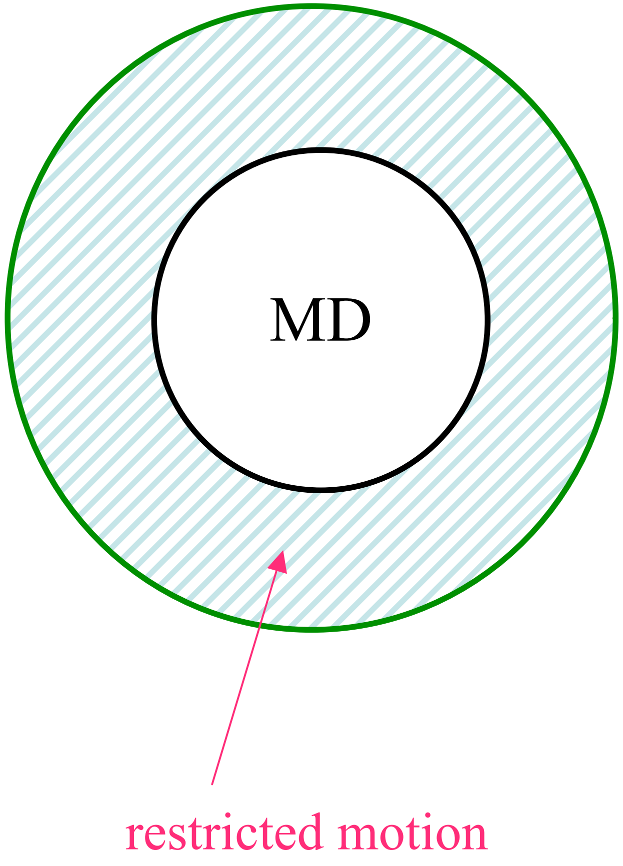
Options:

a) Harmonic position restraining

b) Mean force from outside

[vacuo (attractive + repulsive)]

c) Stochastic force from outside.



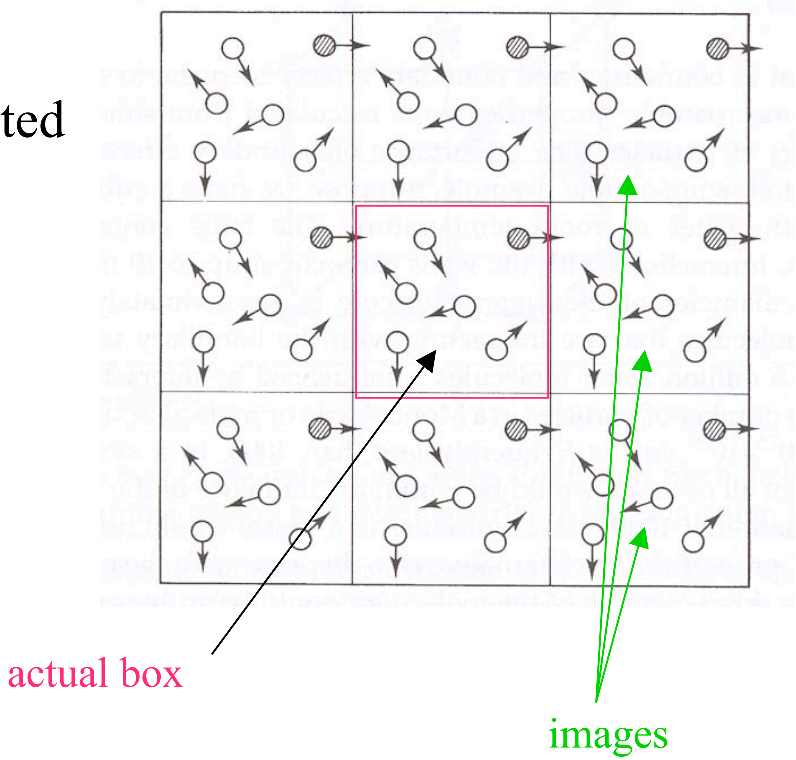
Treatment of Boundaries

3. Periodic:

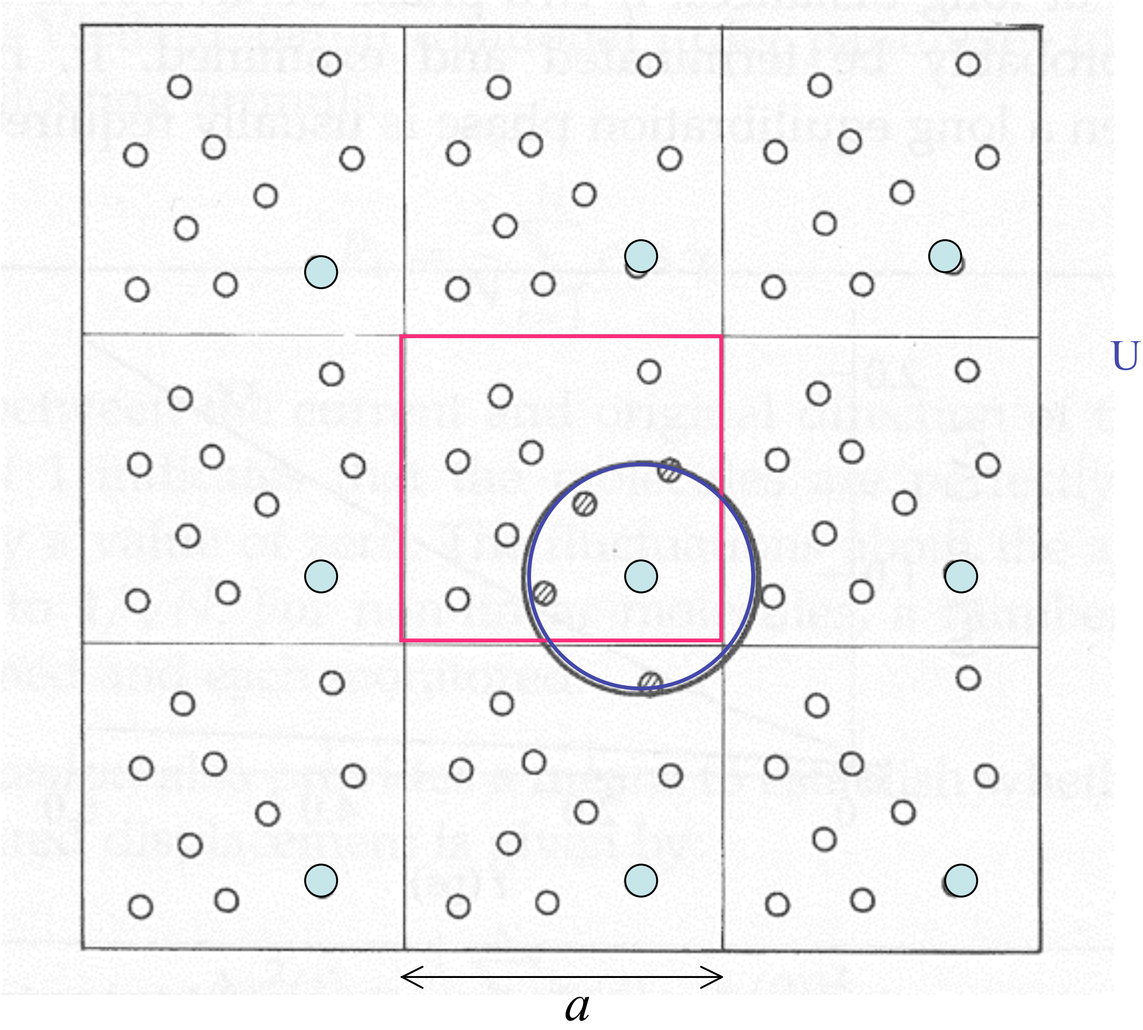
Simulate an infinite periodic system

cube, rectangle, dodecahedron, truncated octahedron

Distortive effects: Anisotropic → cube corners



Minimum Image Convention



Use only nearest image when calculating interaction

cutoff < $a/2$

Molecular Dynamics with Coupling to an External Bath

Standard MD:

independent parameters:
number of particles, N
volume, V
total energy, E_{tot}

conserved quantities:
E_{tot}
$\vec{p}_{tot}$ = total momentum
$\vec{L}_{tot}$ = total angular momentum
*(only in vacuo, **not** with PBC)*

dependent quantities:
temperature = T
pressure = P

Why apply MD at constant T and/or P?

-study system properties as a function of T and P rather than E and V
(normally measure at constant T and P)

-long-range Coulomb force: ~ r-2
use of cut-off radius
} *noise → rising -T*

-study of non-equilibrium systems: P or T gradient.

Couple temperature by scaling the velocities. Couple pressure by scaling the positions.

MD at constant temperature pressure

Pressure

- The pressure fluctuates much more than the energy temperature.
  difference between 2 very large numbers
  virial (attractive forces) and kinetic energy (expansion)

- Isotropic coupling (uniform scaling in x,y,z directions)
- Anisotropic coupling (e.g. for membrane simulations)
- Coupled to temperature coupling.

- **Temperature**
- Flow of kinetic energy weakly coupled parts of systems
- (equipartition not maintained using weak coupling)
- Must separately couple subsystems

What governs the outcome of a MD simulation?

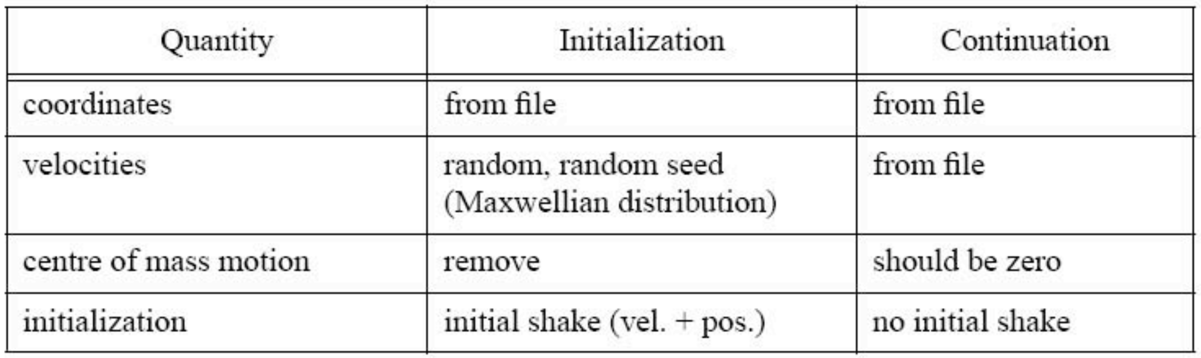
- choice of degrees of freedom
- force field parameters
- treatment of non-bonded interactions
- solvation effects
- boundary conditions
- treatment of temperature and pressure
- integration time step
- starting configuration

Starting an MD Simulation

Input files:

1. description of molecule(s) (topology)
2. starting coordinates (x,y,z)
3. Simulation parameters+ velocities + box?

Simulation parameters:



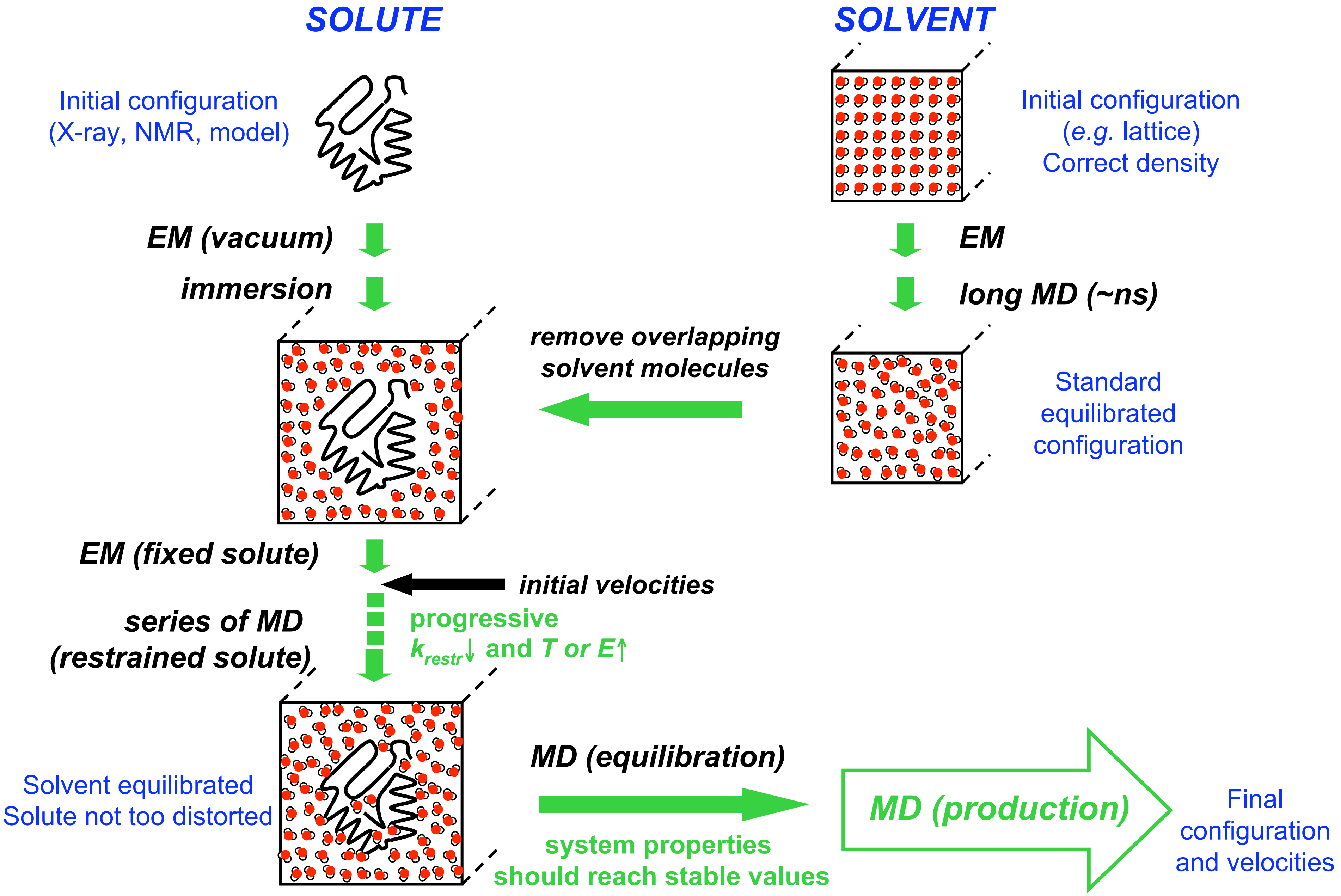
Quantity	Initialization	Continuation
coordinates	from file	from file
velocities	random, random seed (Maxwellian distribution)	from file
centre of mass motion	remove	should be zero
initialization	initial shake (vel. + pos.)	no initial shake

Other choices:

1. Boundary conditions (vacuum, periodic, extended wall)
2. Temperature coupling (yes/no)
3. Pressure coupling (yes/no)
4. Time step
6. Constraints (H-atoms, all atoms)
7. Cut-off (short-range, long-range)
8. Long-range corrections (reaction field)
9. Write (velocities, coordinates, energies?)
10. Special forces (position restraints, distance restraints)

MOLECULAR DYNAMICS (SETUP)

Starting a simulation



SOLUTE

Initial configuration
(X-ray, NMR, model)

EM (vacuum)

immersion

EM (fixed solute)

initial velocities

series of MD
(restrained solute)

progressive
$k_{restr}\downarrow$ and *T or E*$\uparrow$

Solvent equilibrated
Solute not too distorted

SOLVENT

Initial configuration
(*e.g.* lattice)
Correct density

EM

long MD (~ns)

Standard
equilibrated
configuration

remove overlapping
solvent molecules

MD (equilibration)

system properties
should reach stable values

MD (production)

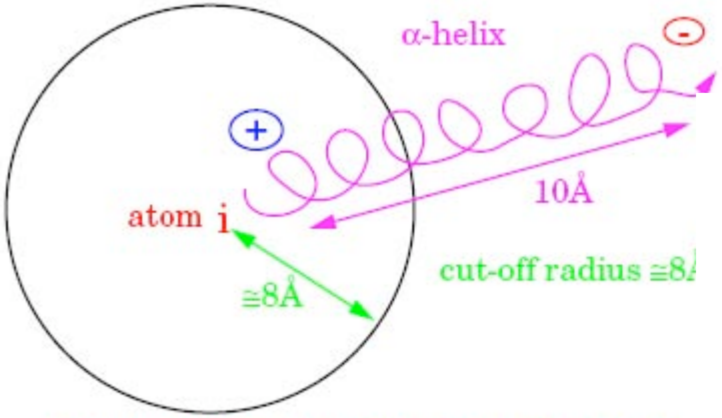
Final
configuration
and velocities

Treatment of Long Range Forces

Treatment of Long Range Forces (cont.)

1. Inclusion of all atoms pairs: computing time ~N^2 (too expensive)

2. Cutoff radii

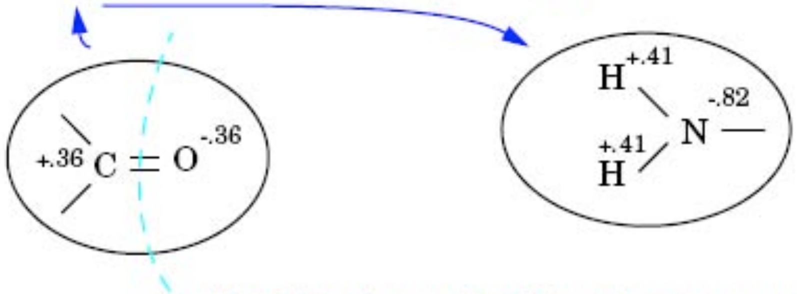


No interactions beyond a distance $r_{ij} > R_c$

Error:
charge-charge r^{-1}
charge-dipole r^{-2}
dipole-dipole r^{-3}
↓ shorter range

Charge group concept:
(effective dipoles)

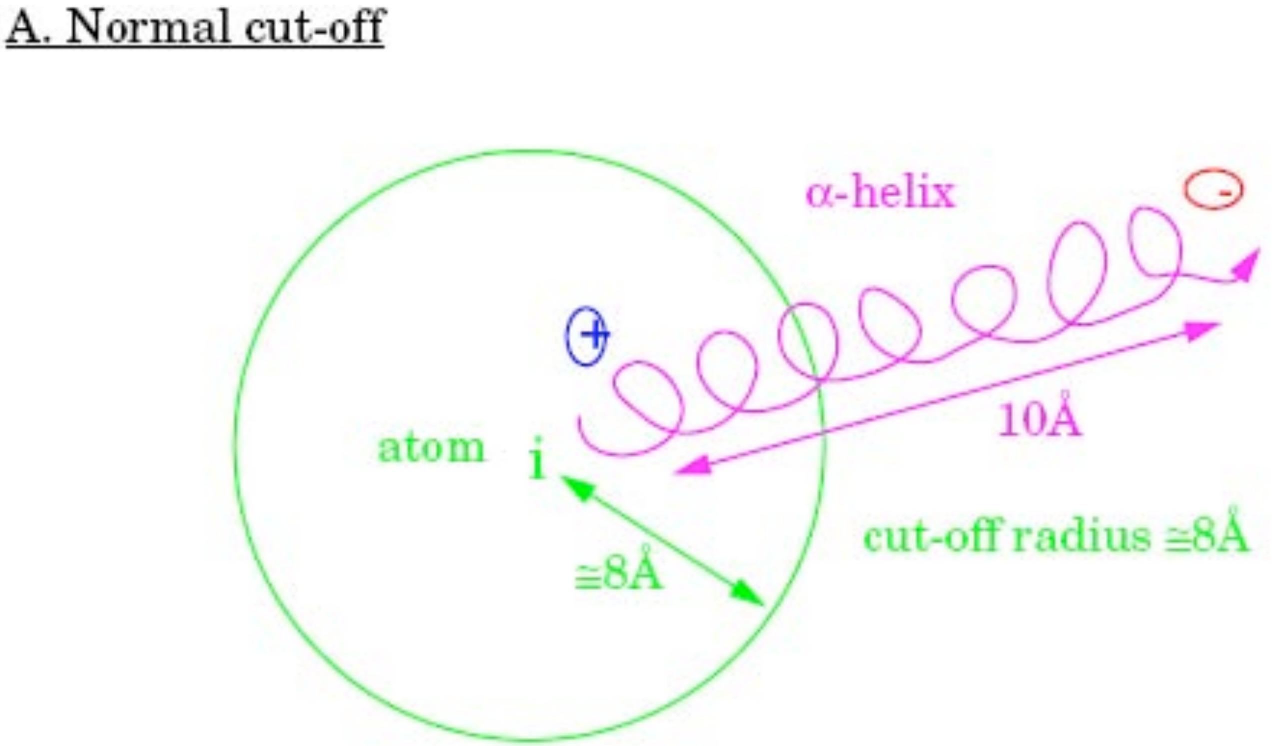
Use neutral groups of atoms: dipole-dipole r^{-3}



cut-off radius does not split a charge group

Treatment of Long Range Forces (cont.)

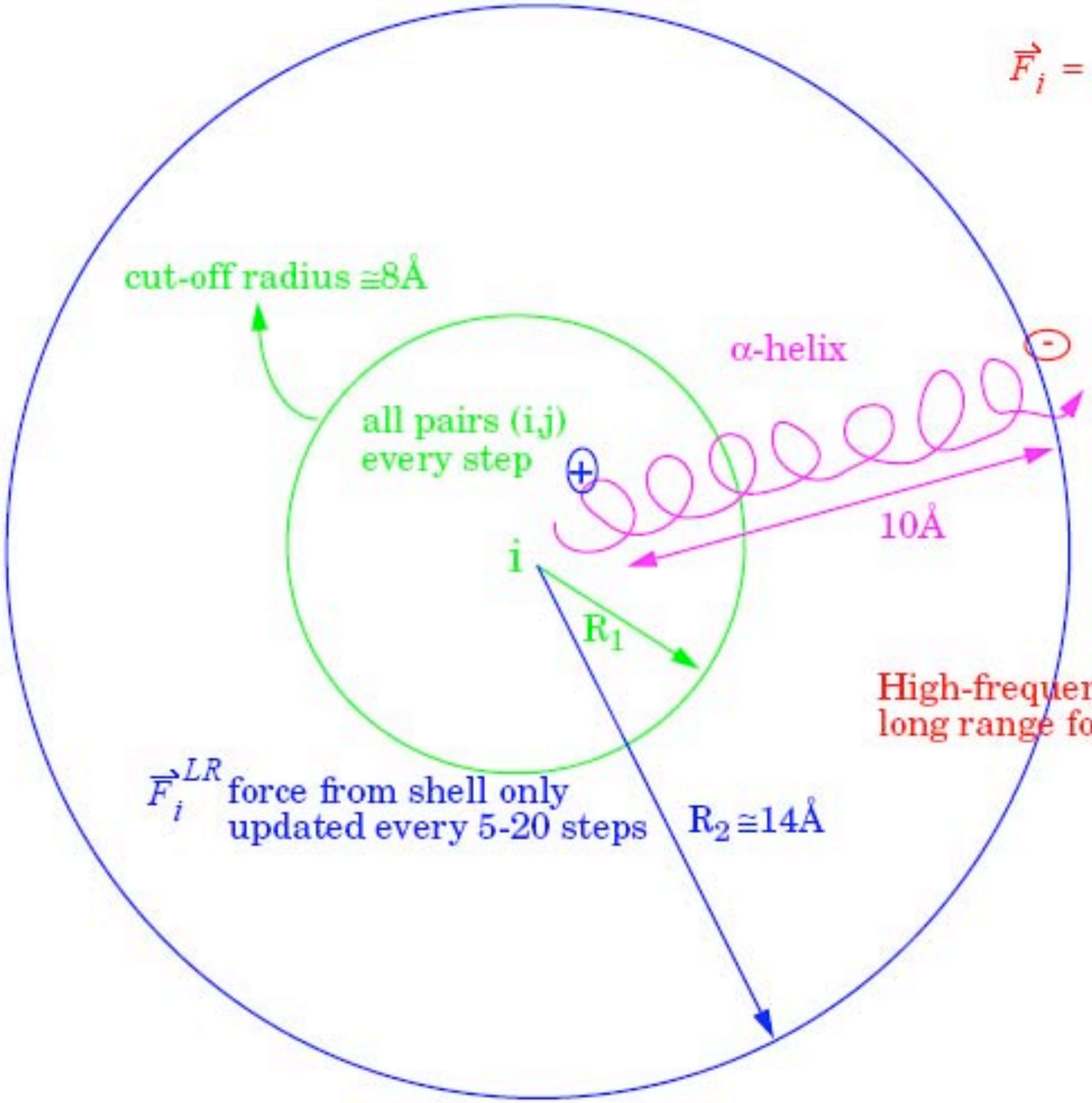
3. Twin Range Cutoff



Treatment of Long Range Forces (cont.)

B. Short range cut-off + semi fixed long range cutoff: Twin range

$$\vec{F}_i = \vec{F}_i^{SR} + \vec{F}_i^{LR}$$



cut-off radius ≅8Å

all pairs (i,j)
every step

α-helix

$10Å$

$\vec{F}_i^{LR}$ force from shell only
updated every 5-20 steps

$R_2 \cong 14Å$

High-frequency components of
long range force neglected