NATIONAL INSTITUTE OF CHEMISTRY

Euro CC Workshop on

Computational and Statistical Methods in Chemistry

Molecular Dynamics Simulation and its Application

Franci Merzel

Laboratory for Molecular Structural Dynamics
National Institute of Chemistry
Ljubljana

October 19 2021

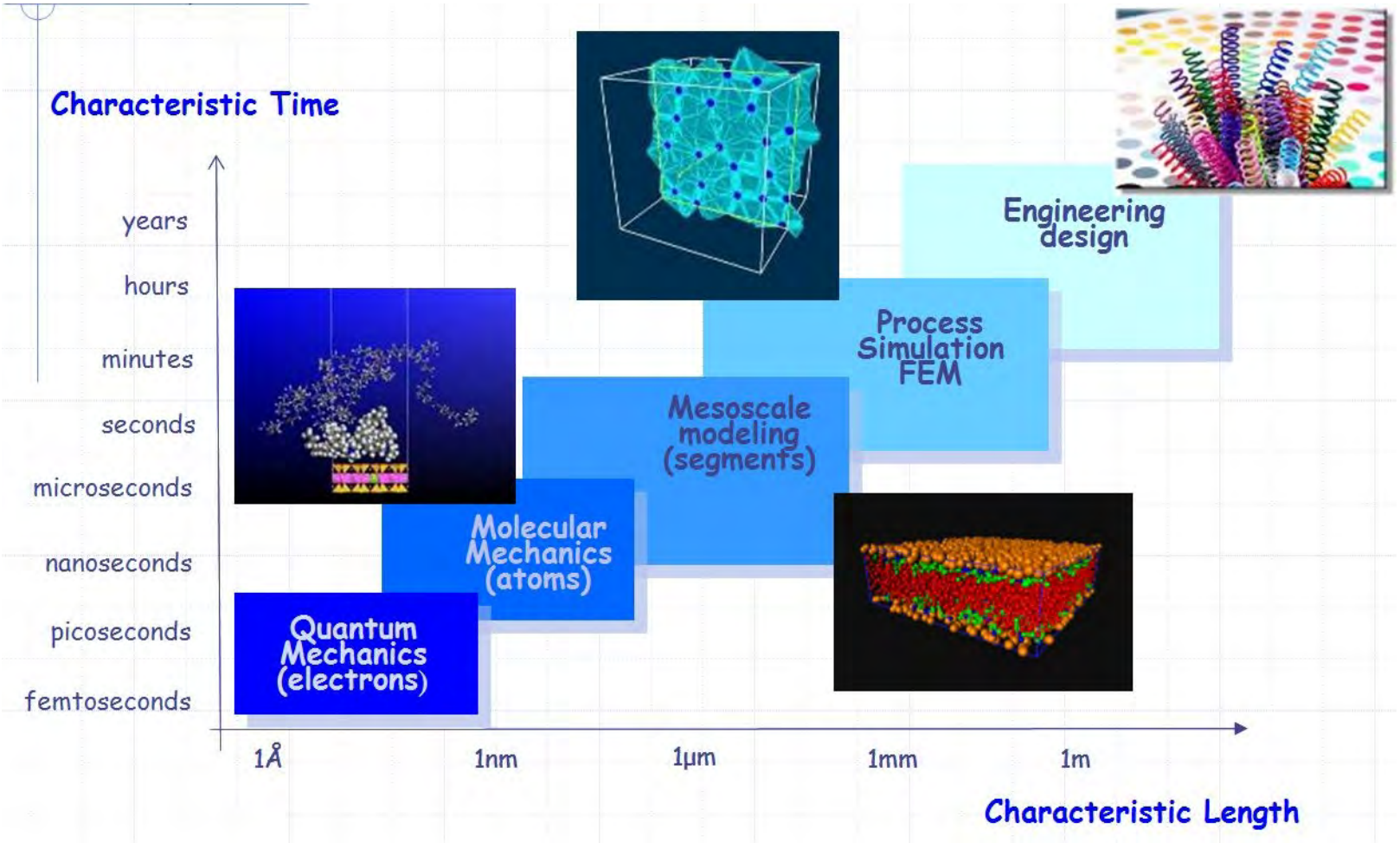
Outline

Introductory theoretical framework of MD simulations

- classical force field MD
- *ab-initio* MD
- coarse-grained MD

Background motivation: **understanding nature**

Modeling at different length- and time- scales
depending on the phenomena we look at:



Macroscale & Mesoscale:
- times > ~ 10^{-6} sec
- lengths > ~ 1000 Å
- coarse graining, DPD, Lattice Boltzmann, FEM

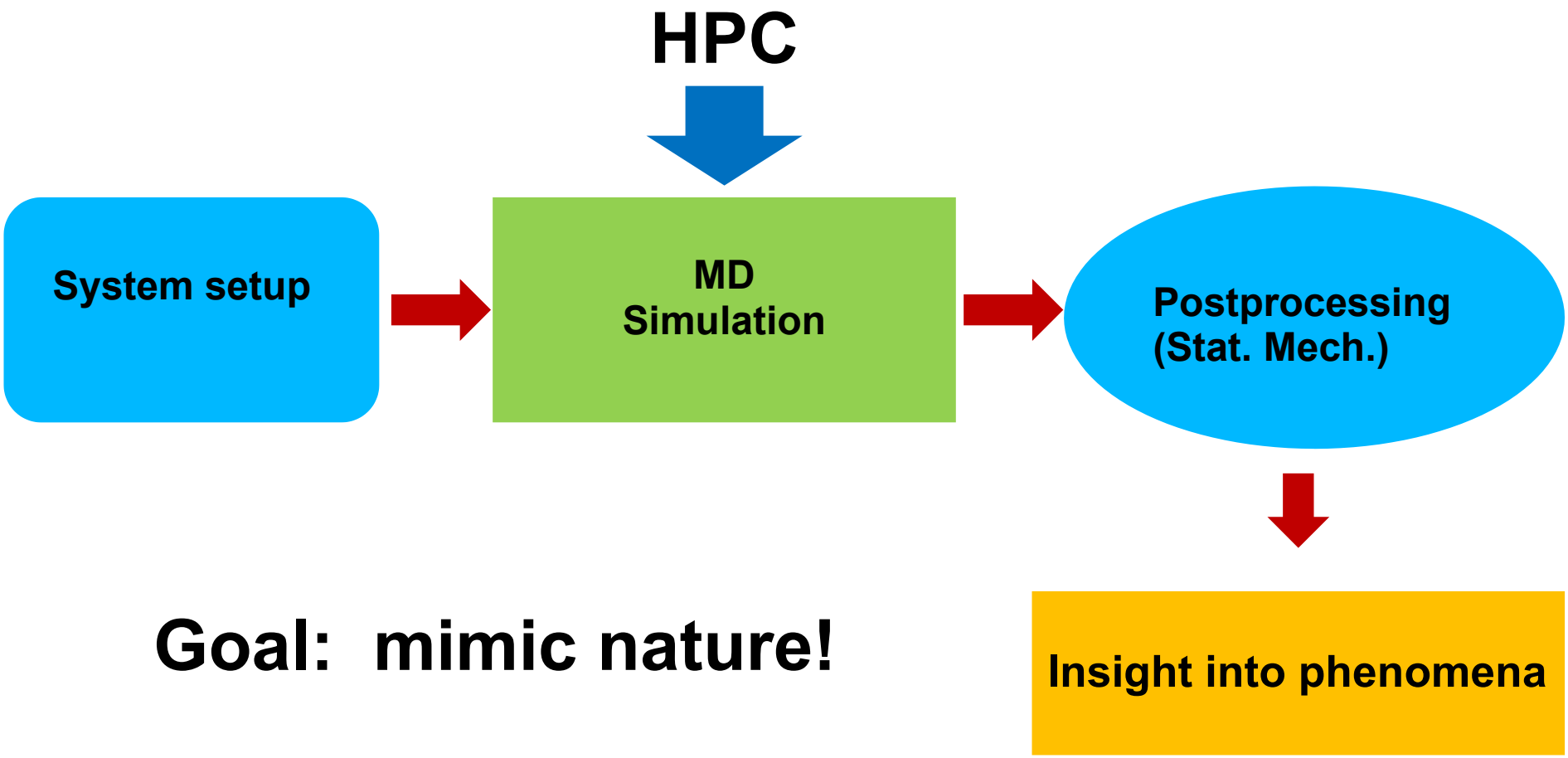
Atomistic scale:
- times ~ 10^{-15}–10^{-6} sec
- lengths ~ 1–100Å
- Molecular Dynamics, Monte Carlo

Understanding the matter
in terms of
"jiggling and wiggling of atoms"

Subatomic scale:
- times ~ 10^{-15}–10^{-9} sec
- lengths ~ 1–10 Å
- electronic structure (DFT),
- ab initio MD, CP, BO

Essential prerequisite for any simulation in available computer power,
lots of computing power**… HPC**

Motivation: simulation chart



Goal: mimic nature!

Validation with experiment

Bottom-line description of atomic behaviour is **quantum mechanics**

Stationnary states of the Schrödinger equation of a molecule

$$H\,\Psi(\boldsymbol{r_1}, \boldsymbol{r_2}, \ldots | \boldsymbol{R_A}, \boldsymbol{R_B}, \ldots) = E\,\Psi(\boldsymbol{r_1}, \boldsymbol{r_2}, \ldots | \boldsymbol{R_A}, \boldsymbol{R_B}, \ldots)$$

Exact Hamiltonian:

$$H = T_n(\boldsymbol{R_A}, \boldsymbol{R_B}, \ldots) + V_{nn}(\boldsymbol{R_A}, \boldsymbol{R_B}, \ldots) + \boldsymbol{H_e}\,(\boldsymbol{r_1}, \boldsymbol{r_2}, \ldots | \boldsymbol{R_A}, \boldsymbol{R_B}, \ldots)$$

Kinetic energy of the n

Electrostatic Repulsion n-n

Hamiltonian of the electrons Kinetic-e +potential-e-e+ **potential e-n**

Because the nuclei mass >> electron mass, we can decouple the electrons and the nuclei

Approximate Hamiltonian:

PES

$$H \approx T_n(\boldsymbol{R_A}, \boldsymbol{R_B}, \ldots) + V_{nn}(\boldsymbol{R_A}, \boldsymbol{R_B}, \ldots) + E_e(\boldsymbol{R_A}, \boldsymbol{R_B}, \ldots)$$

Average e energy for given positions of the n

Modern Quantum Chemistry, A. Szabo & N.S. Ostlund, Dover, 1996

Summary

Because the nuclei mass >> electron mass, we can decouple the electrons and the nuclei

$$H \simeq T_n(\boldsymbol{R}_1, \boldsymbol{R}_2, \dots) + U(\boldsymbol{R}_1, \boldsymbol{R}_2, \dots)$$

Kinetic energy of the nuclei (atoms)

PES depending on the nuclei (atomic) coordinates only and is represented by a **Force-Field**

Quantum case

$$T_n(\boldsymbol{R}_1, \boldsymbol{R}_2, \dots) = \frac{\hbar^2}{2}\left[\frac{\Delta_{R_1}}{M_1} + \frac{\Delta_2}{M_2} + \cdots\right]$$

Equation of motion

$$H\chi(\boldsymbol{R}_1, \boldsymbol{R}_2, \dots; t) = i\hbar\frac{\partial\chi(\boldsymbol{R}_1, \boldsymbol{R}_2, \dots; t)}{\partial t}$$

Solutions : wave-funtion $\chi(\boldsymbol{R}_1, \boldsymbol{R}_2, \dots; t)$

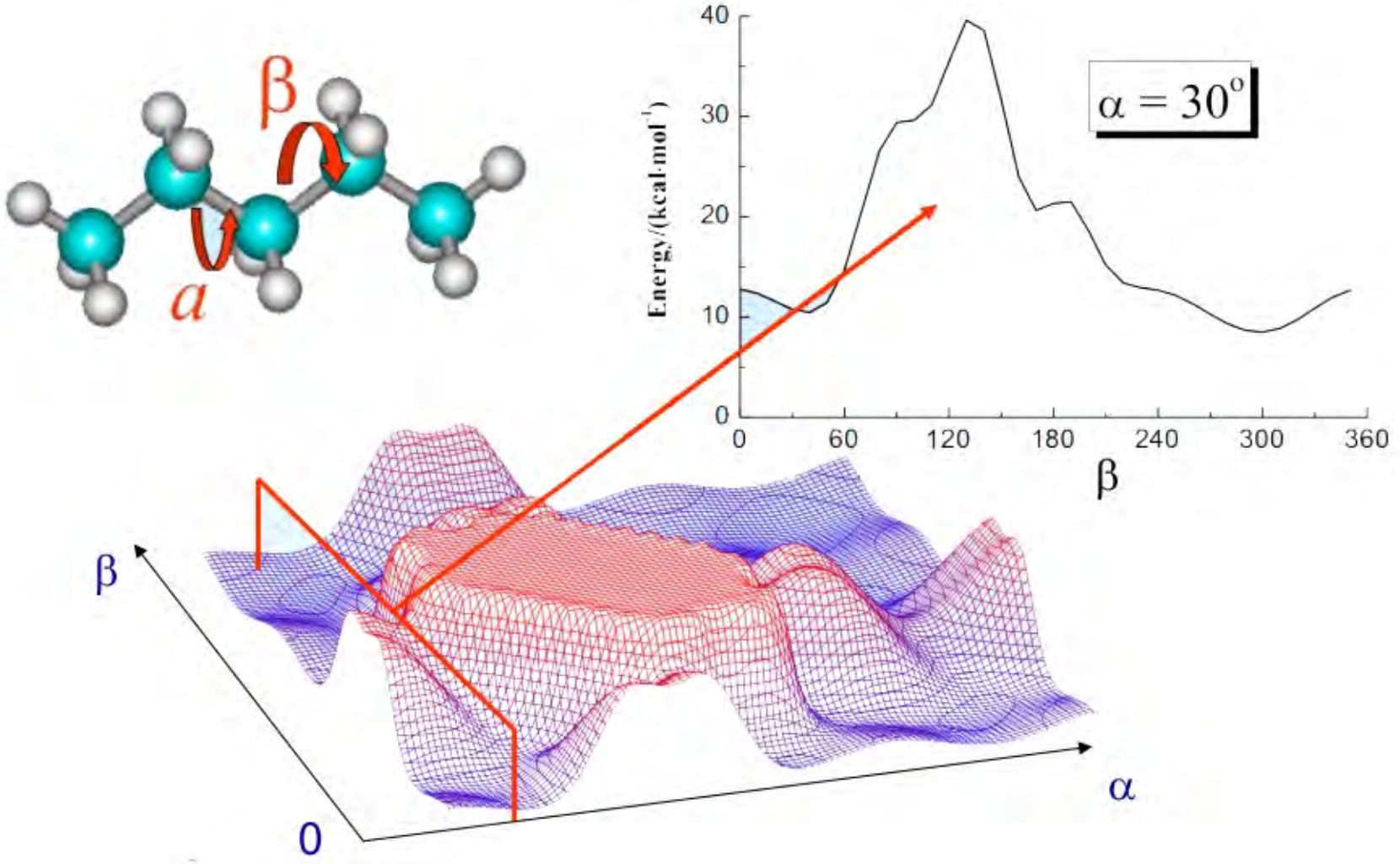
Classical case, MD

$$T_n(\boldsymbol{R}_1, \boldsymbol{R}_2, \dots) = \frac{1}{2}\left[M_1 v_1^2 \quad + M_2 v_2^2 \quad \dots\right]$$

$$-\frac{\partial H}{\partial \boldsymbol{R}_I} = \boldsymbol{F}_I = M_I \dot{v}_I$$

Solutions : trajectory $\{\boldsymbol{R}_1(t), \boldsymbol{R}_2(t) \dots\}$

Potential energy surface (PES) $U(r_1, r_2, \ldots r_N) \rightarrow \underline{U}(\alpha,\beta)$ in terms of internal coordinates α, β



Basic Molecular Dynamics (MD)

Basic lesson from mechanics:
provided we know all the positions and velocities of the point-like particles at a given moment and we know how they interact, than we are able to predict time evolution of such system just by solving Newton's equations of motion.

This is exactly the case of an atomistic system:

$$U = U(\vec{r}_1, \vec{r}_2 \dots \vec{r}_N)$$

$$\vec{F}_i = -\frac{\partial U}{\partial \vec{r}_i} = -\nabla_i U(\vec{r}_1, \vec{r}_2 \dots \vec{r}_N)$$

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

Result of the MD simulation is a **trajectory** of the system: information of positions and velocities of all atoms at every step during the simulation.

Numerical integration using Verlet scheme

For sufficiently short step-size Δt we can write forward and backward Taylor expansion and sum both expressions:

$$\vec{r}_i(t+\Delta t) = \vec{r}_i(t) + \frac{d\vec{r}_i(t)}{dt}\Delta t + \frac{1}{2}\frac{d^2\vec{r}_i(t)}{dt^2}\Delta t^2 + \frac{1}{6}\frac{d^3\vec{r}_i(t)}{dt^3}\Delta t^3 + \mathcal{O}(\Delta t^4)$$

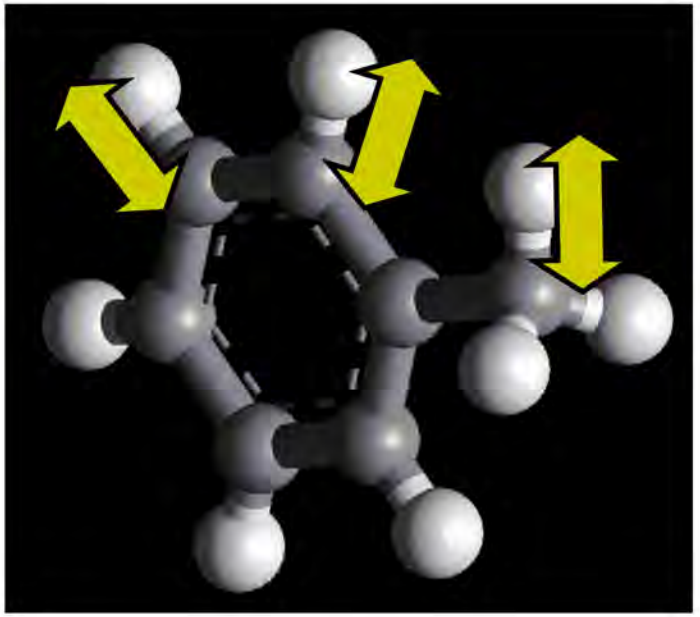
$$\vec{r}_i(t-\Delta t) = \vec{r}_i(t) - \frac{d\vec{r}_i(t)}{dt}\Delta t + \frac{1}{2}\frac{d^2\vec{r}_i(t)}{dt^2}\Delta t^2 - \frac{1}{6}\frac{d^3\vec{r}_i(t)}{dt^3}\Delta t^3 + \mathcal{O}(\Delta t^4)$$

We obtain “coordinate” Verlet algorithm:

$$\vec{r}_i(t+\Delta t) = 2\vec{r}_i(t) - \vec{r}_i(t-\Delta t) + \frac{\vec{F}_i(t)}{m_i}\Delta t^2 + \mathcal{O}(\Delta t^4)$$

The 1 fs Time Barrier

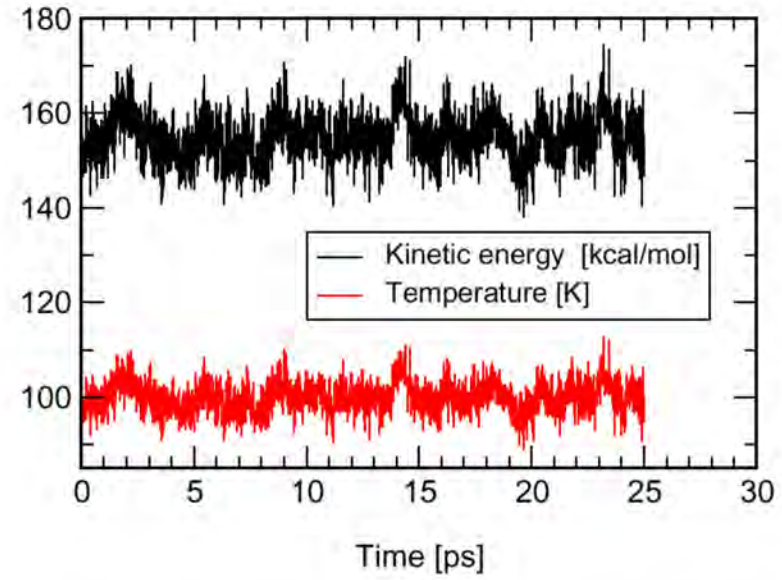
- Dynamics simulations are limited by the highest frequency vibration
- Ideally the timestep should be 1/10 highest frequency
- In most cases C-H bond stretching (10^{-14} s) is the fastest mode

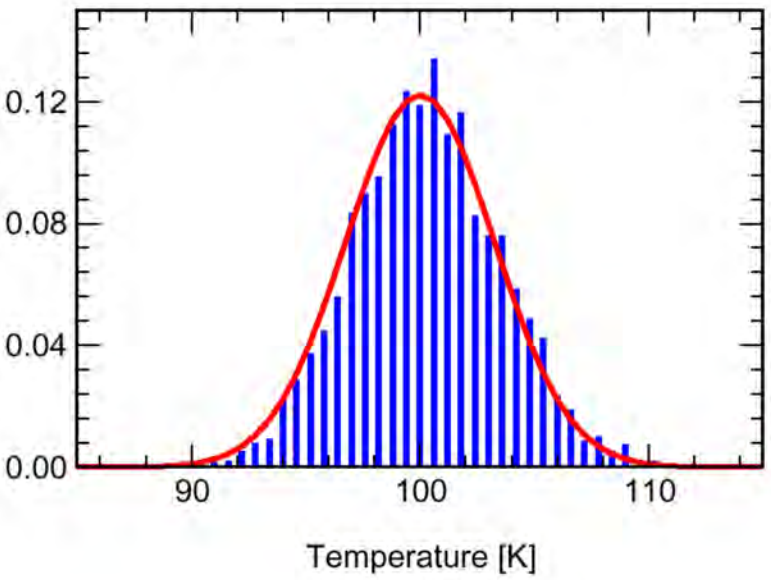


SPEED LIMIT

1 fs

Relationship to thermodynamic quantities (Boltzmann)





Definition of Temperature

$$\langle \sum_j \frac{1}{2} m_j v_j^2 \rangle = \frac{3}{2} N k_B T$$

$$T = \frac{2}{3N k_B} \langle \sum_j \frac{1}{2} m_j v_j^2 \rangle$$

Statistical mechanics provides link between calculated trajectories and thermodynamic quantities.

Our main goal in molecular simulations is to compute statistical averages (i.e. ensemble averages) of observables.

In that respect MD uses the ergodic hypothesis:

$$\langle A \rangle_{trajectory} = \langle A \rangle_{ensemble}$$

The very accuracy of trajectories is not as important as its long term stability.

Simulations should provide correct physical ensembles!!

MD in various Ensembles

Equations of motion are energy conserving
NVE (microcanonical) ensemble

Dynamics can be modified to yield other ensembles:

NVT: canonical ensemble

NPT: isothermal – isobaric ensamble

μPT: Gibbs ensemble, μVT grand-canonical

Often achieved via changing the equations of motions (i.e. barostats, thermostats,…)

Setting up the force field

FORCE FIELD (FF) = ENSEMBLE OF PARAMETERS FOR A MODEL PES

ALL-ATOM FF = FUNCTION OF ALL ATOMIC COORDINATES (THOUSANDS IN MACROMOLECULE !)

APPROXIMATION

$U(\boldsymbol{R}_1, \boldsymbol{R}_2, \ldots \boldsymbol{R}_N) \cong$ SUM OF POTENTIALS BETWEEN 2 TO 4 ATOMS

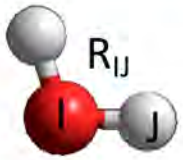
$$U(\boldsymbol{R}_1, \boldsymbol{R}_2, \ldots \boldsymbol{R}_N) = \sum_{I,J}^{N(N-1)/2} U_2(\boldsymbol{R}_I, \boldsymbol{R}_J)$$

$$\sum_{I,J,K}^{N(N-1)(N-2)/6} U_3(\boldsymbol{R}_I, \boldsymbol{R}_J, \boldsymbol{R}_K) + \sum_{I,J,K,L}^{N(N-1)(N-2)(N-3)/24} U_4(\boldsymbol{R}_I, \boldsymbol{R}_J, \boldsymbol{R}_K, \boldsymbol{R}_L)$$

TYPICAL POTENTIAL IN AN ALL-ATOM FORCE FIELD FOR A MACROMOLECULE (PROTEIN)

Covalent bonds (harmonic potentials):

$$U_2(\boldsymbol{R}_I, \boldsymbol{R}_J) = \frac{1}{2} K_{IJ} \left(R_{IJ} - {R_{IJ}}^0 \right)^2$$



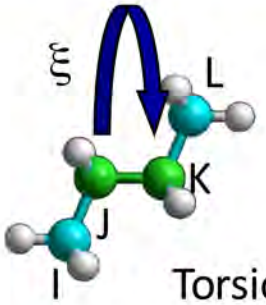
Stretching

$$U_3(\boldsymbol{R}_I, \boldsymbol{R}_J, \boldsymbol{R}_K) = \frac{1}{2} C_{IJK} \left(\theta_{IJK} - {\theta_{IJK}}^0 \right)^2$$

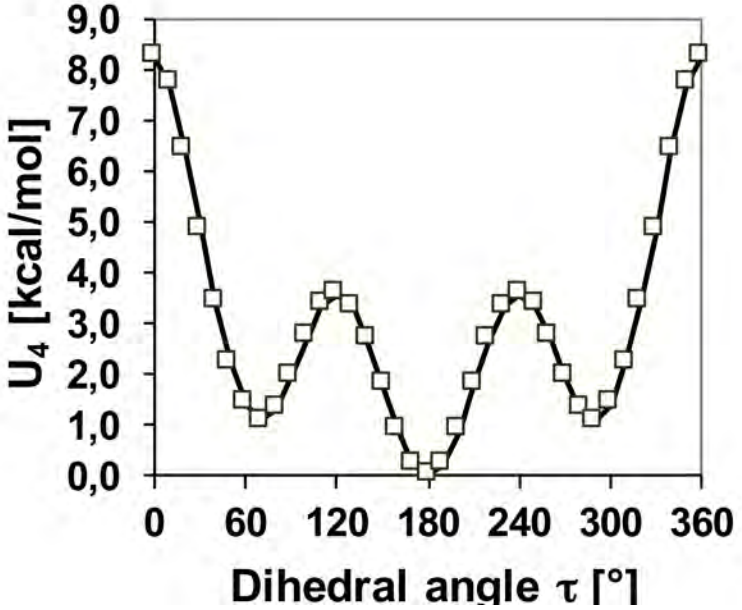
Bond bending

Covalent bonds (anharmonic potentials):

$$U_4(\boldsymbol{R}_I, \boldsymbol{R}_J, \boldsymbol{R}_K, \boldsymbol{R}_L) = \frac{1}{2} V_{IJKL} \left[1 + \cos(n\xi_{IJKL} - {\xi_{IJKL}}^0) \right]$$



Torsional



Non-bonding interaction

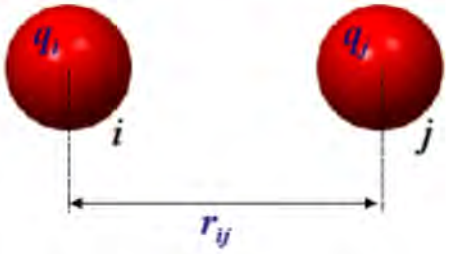
$$U_2(\boldsymbol{R}_I, \boldsymbol{R}_J) =$$

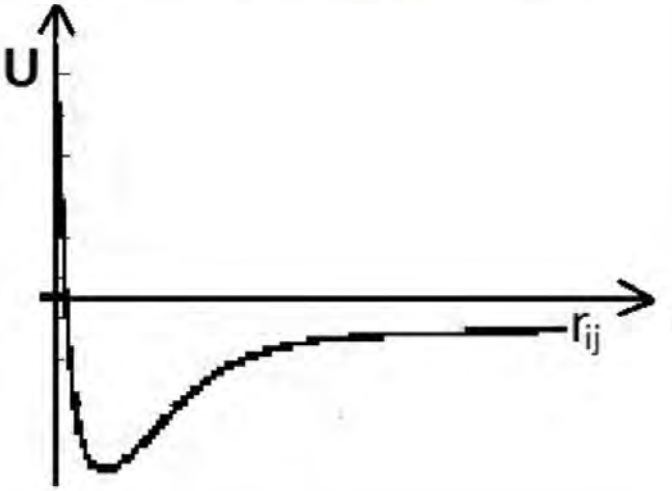
$$4\varepsilon_{IJ}\left[\frac{\sigma_{IJ}{}^{12}}{R_{IJ}{}^{12}} - \frac{\sigma_{IJ}{}^{6}}{R_{IJ}{}^{6}}\right]$$

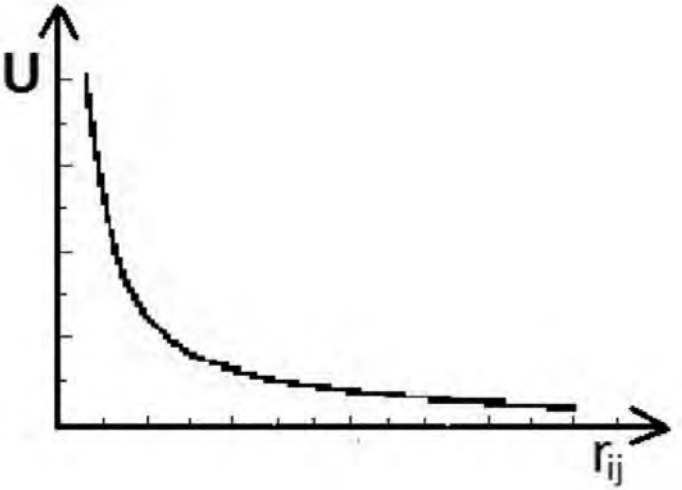
$$+\frac{q_I q_J}{R_{IJ}}$$

Van der Waals interaction
Steric - short range

electrostatic interaction
long range!!!







Partial charges ql usually obtained by
the population analysis

Special treatment of electrostatic
interaction required

Cutoffs –> shifting or switching potentials

Ewald summation,
Particle mesh Ewald

How to obtain the FF parameters?

• fitting experimental data: vibrational spectra, density, g(r), diffusion, heat of vaporization...

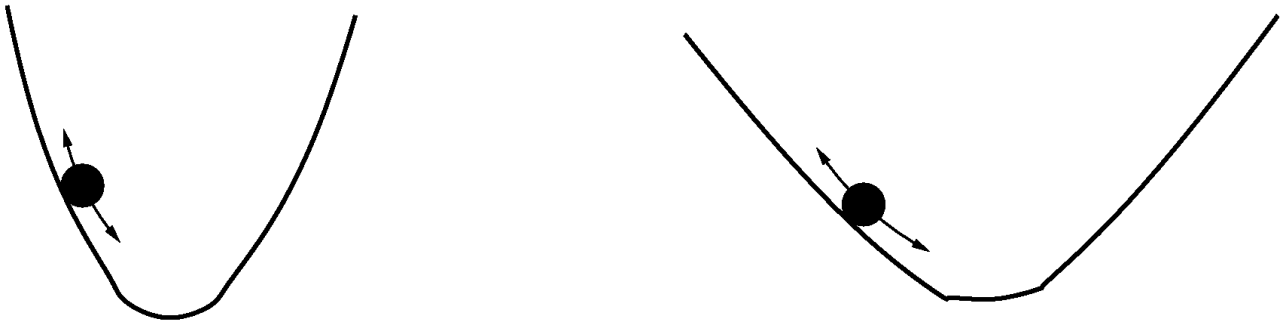
• using QM calculations to calculate some interaction parameters!

• in general, FF work best for the situation for which they were parametrized; it is an extra check required to prove their transferability

• combining FF parameters is dangerous and non-trivial, it often needs re-parametrization

bottom up vs. top down approaches

What can we learn from vibrations?



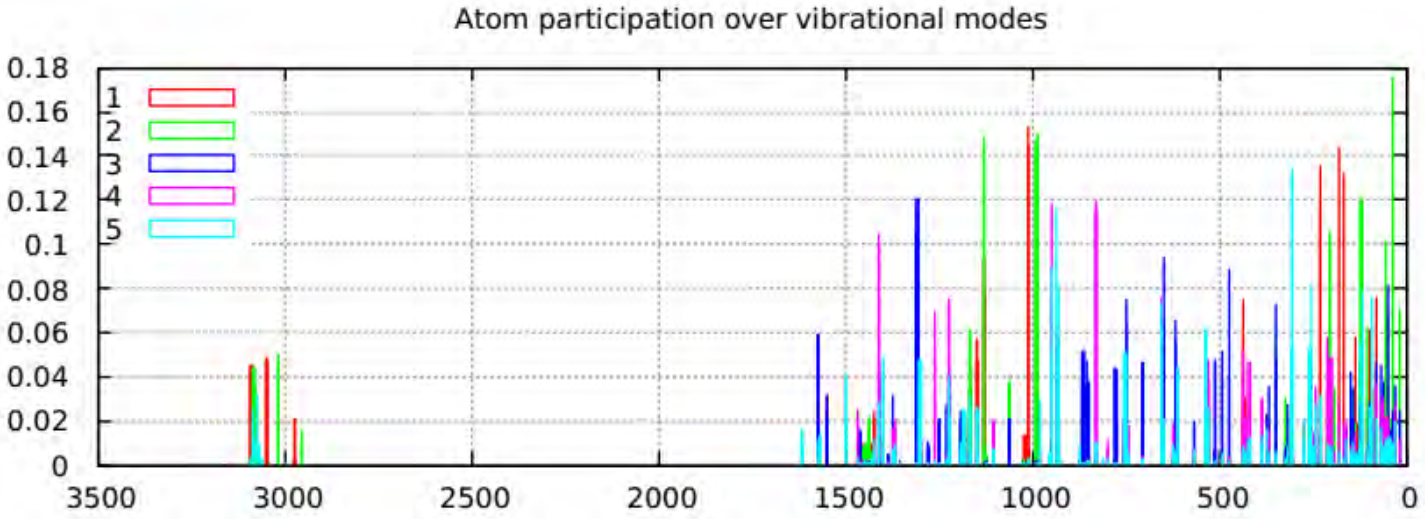
$\omega_1 > \omega_2$

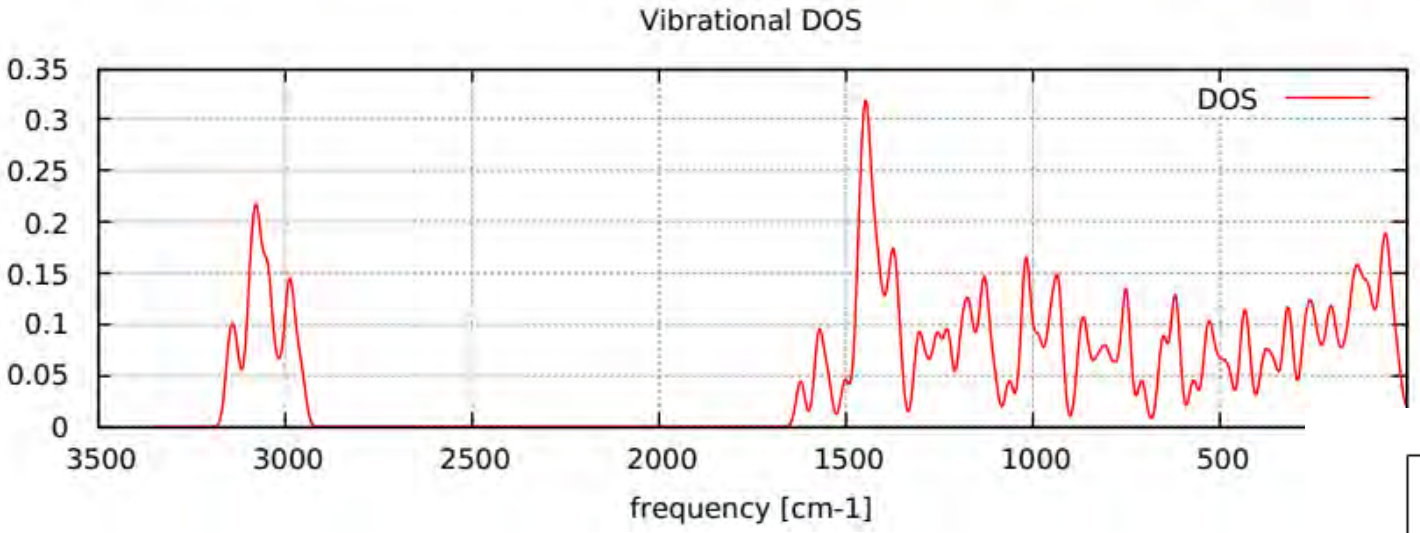
Vibrations: probe the potential energy surface – of all atoms in all directions. → they measure the forces holding the atoms together.

Analytically the most elegant way of describing vibrations of an interacting system of particles is by means of **normal modes**.

This is possible within the **harmonic approximation** which provides a description of internal vibrations in terms of decoupled collective displacements called normal modes.

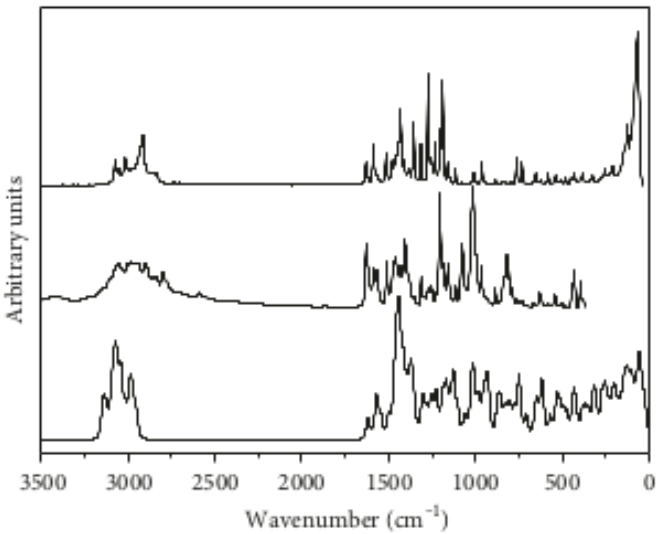
Vibrational spectra: "natural" control for the quality of the FF





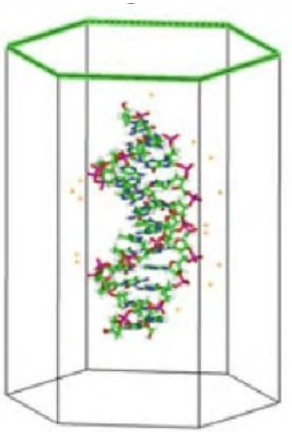
DFT calculation of vibrational spectrum

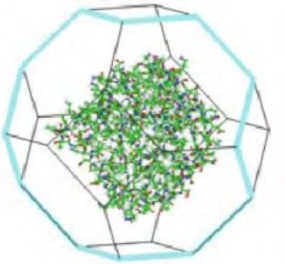
Comparison with experiment

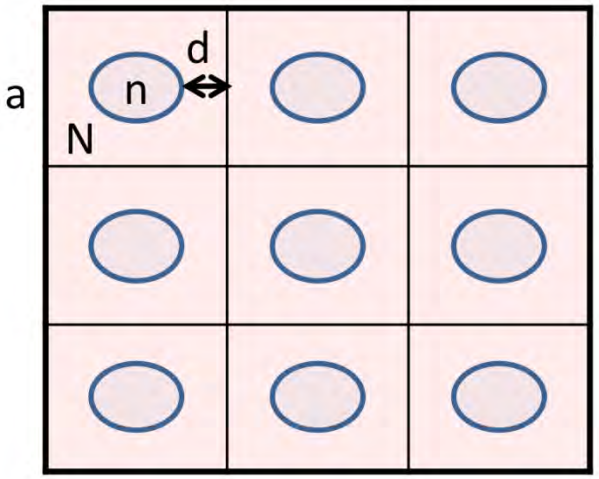


Performing a typical atomistic simulation

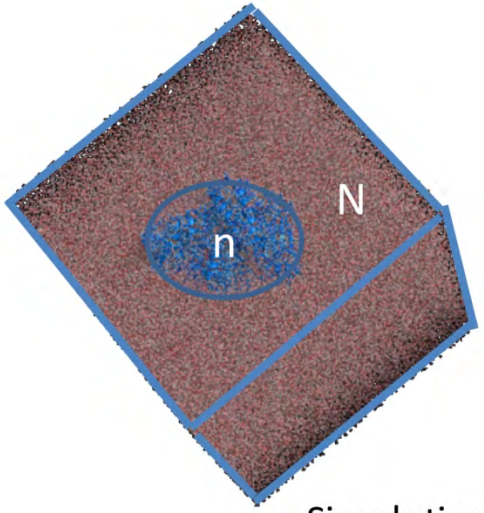
We need to:
- provide initial structure (experimental): n atoms
- define a simulation box (assuming Periodic Boundary Conditions (PBC))
- add solvent (explicit): N atoms







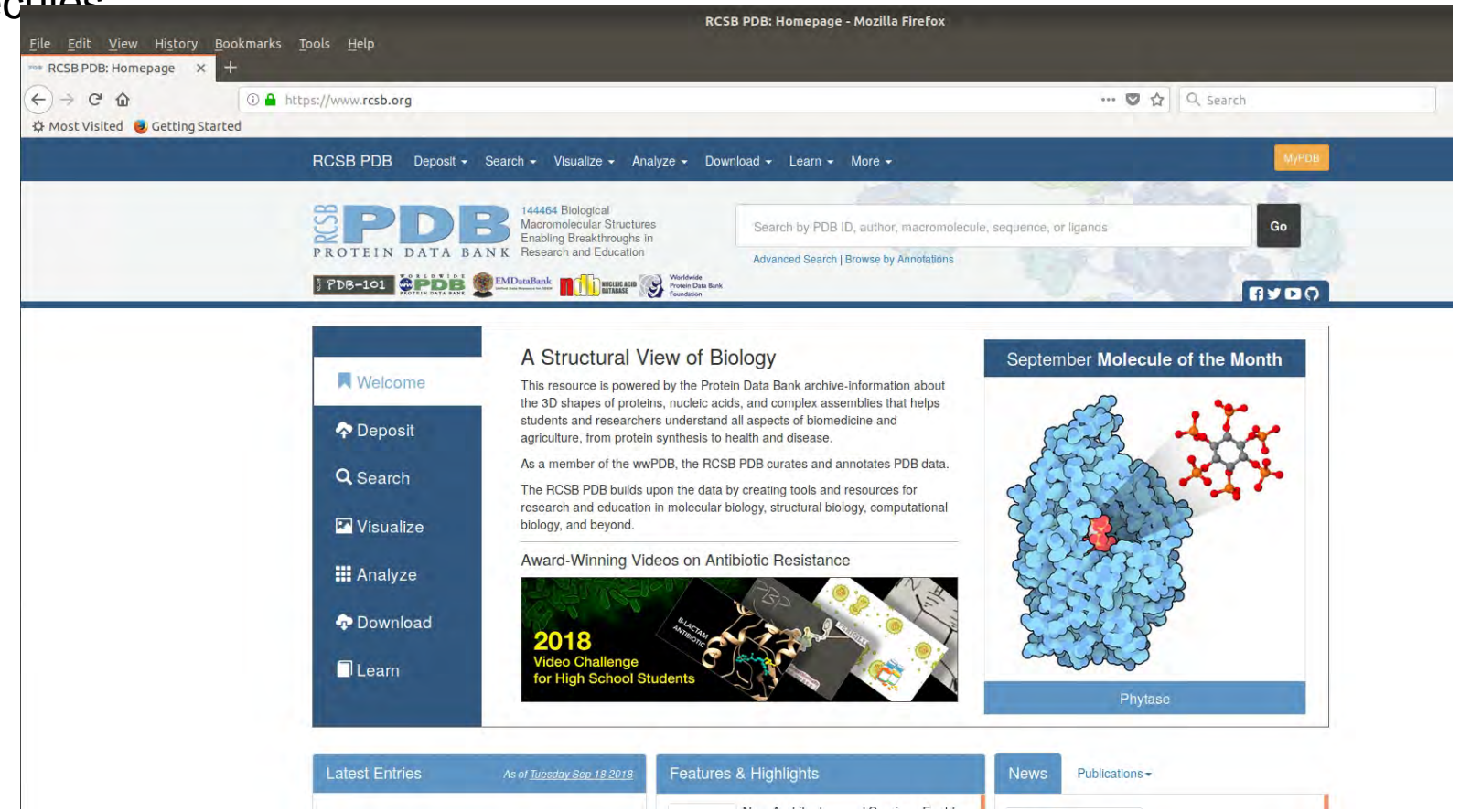
Typical values:
n ~ 5000
N ~ 100000
a ~ 10nm
d ~10A



Simulation cell

Required input 1: structure (initial coordinates) and velocities

Structure ← Protein Data Bank for complex biomolecules,
first principles geometry optimizations for small molecules



Velocities set according to the Maxwell distribution law for a given temperature

Required input 2: topology and parameters (FF)

Topology files contain:

- atom types are assigned to identify different elements and different molecular orbital environments
- charges are assigned to each atom
- connectivities between atoms are established

Parameter files contain:

- force constants necessary to describe the bond energy, angle energy, torsion energy, nonbonded interactions (van der Waals and electrostatics)
- suggested parameters for setting up the energy calculations

coordinates

```
...
...
  317    22 GLN  HG2    -8.84400 -12.92800  20.60500 KRAS 22      0.00000
  318    22 GLN  CD     -7.90200 -12.94900  22.53700 KRAS 22      0.00000
  319    22 GLN  OE1    -6.70400 -12.68600  22.75200 KRAS 22      0.00000
  320    22 GLN  NE2    -8.84300 -12.60300  23.35400 KRAS 22      0.00000
  321    22 GLN  HE21   -9.79500 -12.82100  23.14100 KRAS 22      0.00000
  322    22 GLN  HE22   -8.61800 -12.11700  24.19900 KRAS 22      0.00000
  323    22 GLN  C      -5.04400 -14.31700  19.81200 KRAS 22      0.00000
  324    22 GLN  O      -4.12800 -13.86400  20.49300 KRAS 22      0.00000
  325    23 LEU  N      -4.93000 -15.42400  19.08500 KRAS 23      0.00000
  326    23 LEU  HN     -5.67700 -15.79500  18.53800 KRAS 23      0.00000
  327    23 LEU  CA     -3.67500 -16.14800  19.06500 KRAS 23      0.00000
  328    23 LEU  HA     -3.40500 -16.38500  20.08300 KRAS 23      0.00000
  329    23 LEU  CB     -3.82500 -17.43100  18.23600 KRAS 23      0.00000
  330    23 LEU  HB1    -3.86900 -17.15500  17.16100 KRAS 23      0.00000
  331    23 LEU  HB2    -4.77800 -17.92600  18.51900 KRAS 23      0.00000
  332    23 LEU  CG     -2.68400 -18.42200  18.43700 KRAS 23      0.00000
  333    23 LEU  HG     -1.72400 -17.86900  18.52100 KRAS 23      0.00000
  334    23 LEU  CD1    -2.91600 -19.23800  19.70900 KRAS 23      0.00000
...
...
```

parameters

```
...
BONDS
!
!V(bond) = Kb(b - b0)**2
!Kb: kcal/mole/A**2
!b0: A
!atom type Kb          b0
!
NH2  CT1   240.000     1.4550
CT2  CD    200.000     1.5220
HA1  CC    317.130     1.1000

...

ANGLES
!
!V(angle) = Ktheta(Theta - Theta0)**2
!Ktheta: kcal/mole/rad**2
!Theta0: degrees
!atom types     Ktheta    Theta0   Kub     S0
!
H    NH2  CT1   50.000    111.00              ! From LSN HC-NH2-CT2
NH2  CT1  CT1   67.700    110.00              ! From LSN NH2-CT2-CT2
CT1  CD   OH1   55.000    110.50              ! From ASPP CT2-CD-OH1

...

DIHEDRALS
!
!V(dihedral) = Kchi(1 + cos(n(chi) - delta))
!Kchi: kcal/mole
!n: multiplicity
!delta: degrees
!atom types             Kchi    n   delta
HP   CA   NY   CPT      2.6000  2   180.00 ! adm jr., 12/30/91, for jwk
NY   CPT  CA   CA       3.0000  2   180.00 ! atm, methylindole, 1/17/04, r6r5

...

!V(Lennard-Jones) = Eps,i,j[(Rmin,i,j/ri,j)**12 - 2(Rmin,i,j/ri,j)**6]
!epsilon: kcal/mole, Eps,i,j = sqrt(eps,i * eps,j)
!Rmin/2: A, Rmin,i,j = Rmin/2,i + Rmin/2,j
!atom  ignored    epsilon      Rmin/2
CA     0.000000  -0.070000     1.992400
HR1    0.000000  -0.046000     0.900000

...
```

Input file

```
...
open read unit 1 card name top_all27_prot_na.rtf
read rtf card unit 1

open read unit 1 card name par_all27_prot_na.prm
read param card unit 1

open read unit 1 card name sys.cor
read coor card unit 1

....

!unit cell
set xyax   @ax
set zax    @az

crystal define tetr @xyax @xyax @zax 90.0 90.0 90.0
crystal build cutoff 14.0
image byres sele segi W .or. segi WAT end

....

open write file unit 2 name trajectory.dcd

dyna cpt start nstep 10000000 timestep 0.001 -
    iprfrq 10000 ihtfrq 0 ieqfrq 0 -
    iuncrd 2 iunrea -1 iunvel -1 kunit -1 -
    nprint 10 nsavc 1000 nsavv 0 ihbfrq 0 isvfrq 0 -
    iasors 1 iasvel 1 iscvel 0 ichecw 0 -
    twindh 100.0 twindl -100.0 ntrfrq 10000 -
    pcons pint pmass 500.0 pref 1.0 pgamma 25.0 -
    hoover reft 300.0 tmass 2000.0 tbath 300.0 -
    firstt 300.0 finalt 300.0 teminc 0.0
...
```

topology

```
...
...
RESI LEU          0.00
GROUP
ATOM N    NH1    -0.47  !     |        HD11 HD12
ATOM HN   H       0.31  !  HN-N          | /
ATOM CA   CT1     0.07  !     |   HB1   CD1--HD13
ATOM HA   HB1     0.09  !     |   |    /
GROUP                   !  HA-CA--CB--CG-HG
ATOM CB   CT2    -0.18  !     |   |    \
ATOM HB1  HA2     0.09  !     |   HB2   CD2--HD23
ATOM HB2  HA2     0.09  !   O=C          | \
GROUP                   !     |        HD21 HD22
ATOM CG   CT1    -0.09
ATOM HG   HA1     0.09
GROUP
ATOM CD1  CT3    -0.27
ATOM HD11 HA3     0.09
ATOM HD12 HA3     0.09
ATOM HD13 HA3     0.09
GROUP
ATOM CD2  CT3    -0.27
ATOM HD21 HA3     0.09
ATOM HD22 HA3     0.09
ATOM HD23 HA3     0.09
GROUP
ATOM C    C       0.51
ATOM O    O      -0.51
BOND CB  CA   CG  CB   CD1 CG   CD2 CG
BOND N   HN   N   CA    C   CA   C +N
BOND CA  HA   CB  HB1  CB  HB2  CG  HG   CD1 HD11
...
...
```

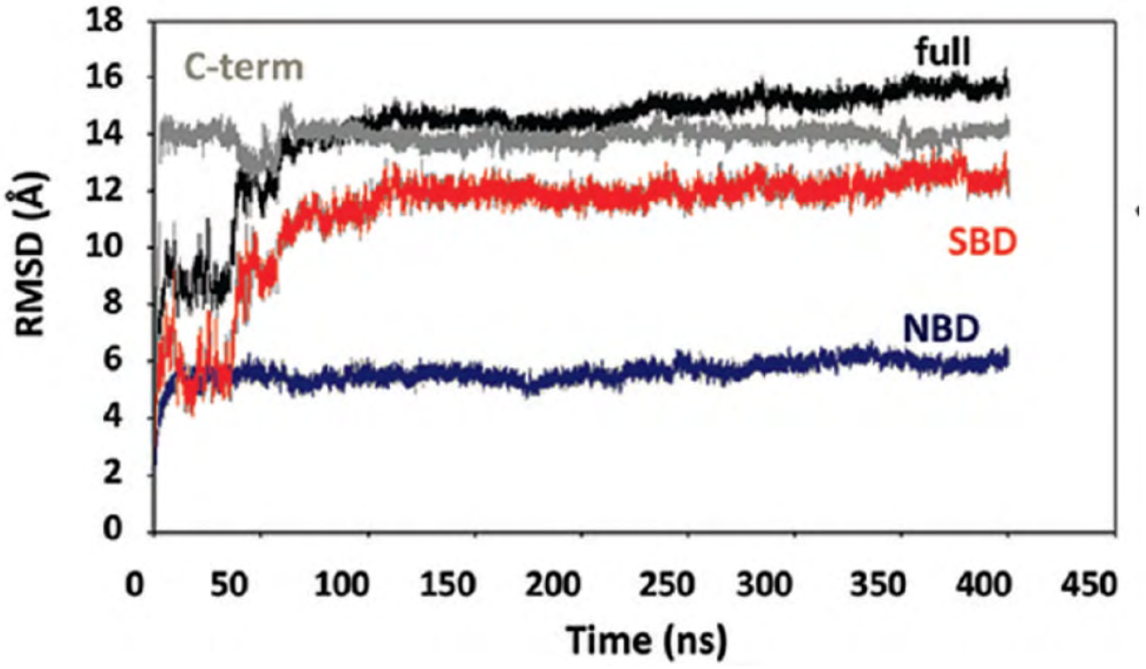
Typical steps in performing MD simulations:

- 1. Prepare molecule
  - Read in pdb and psf file
- 2. Minimization
  - Reconcile observed structure with force field used (T = 0)
- 3. Heating
  - Raise temperature of the system
- 4. Equilibration
  - Ensure system is stable
- 5. Dynamics
  - Simulate under desired conditions (NVE, NpT, etc)
  - Collect your data
- 6. Analysis
  - Collect your data
  - Evaluate observables (macroscopic level properties)
  - Or relate to single molecule experiments

Post-processing

Task 1: Checking whether the system has reached TD equilibrium, i.e. trajectory is stable

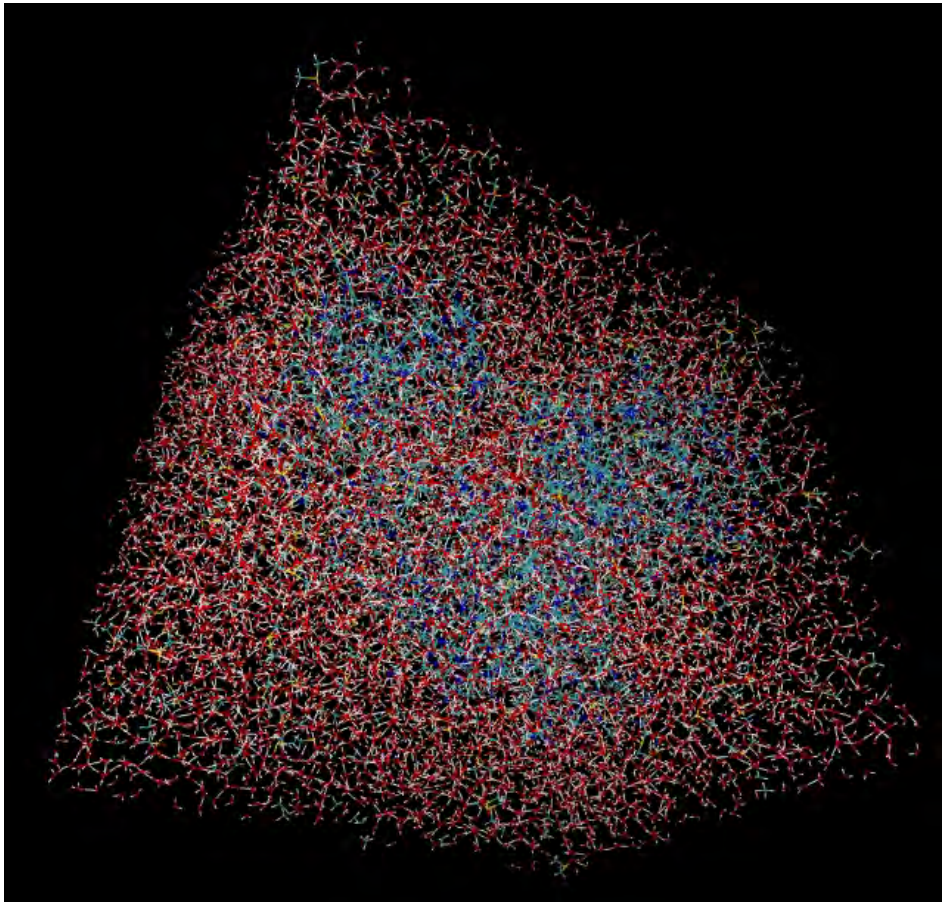
Monitoring of convergence, by using RMSD and Rg

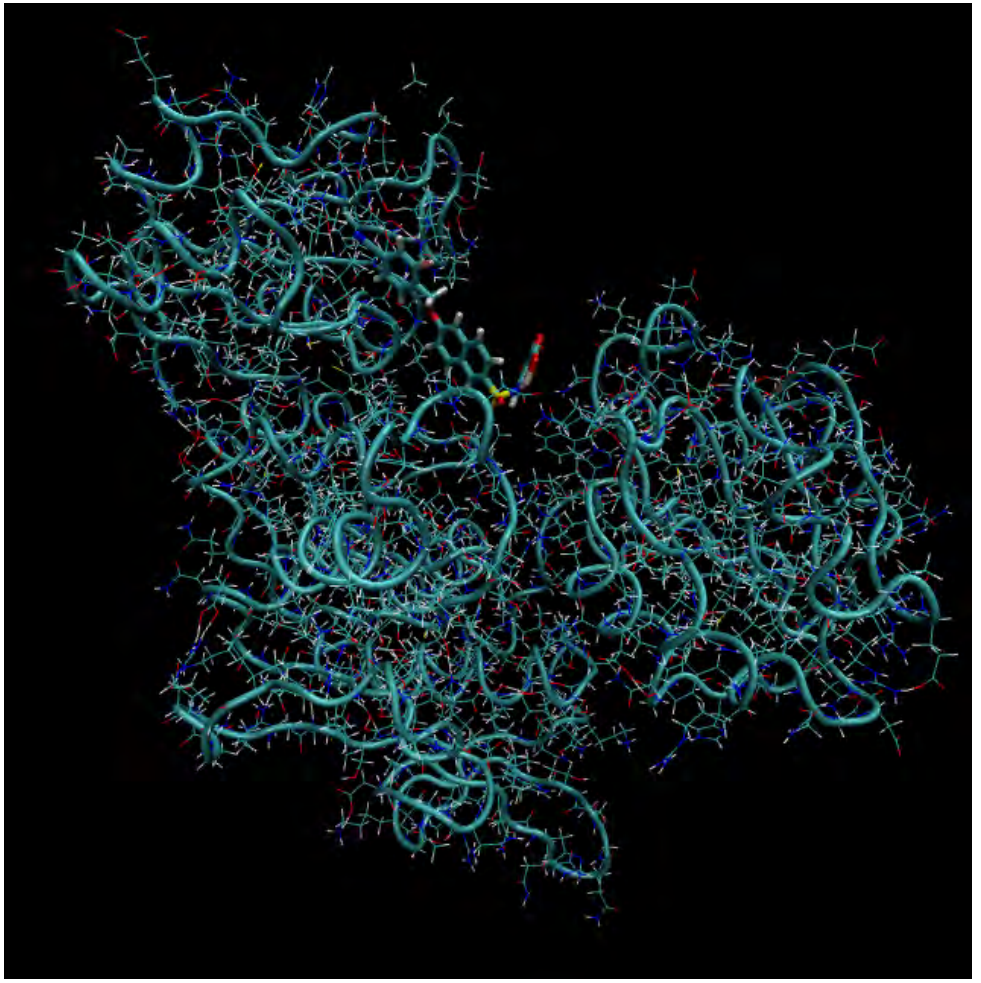


$$\text{RMSD(t)}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\vec{r}(t)-\vec{r}(0))^2}$$

$$R_g = \frac{1}{2N^2}\sum_{i,j}^{N}(\vec{r}_i-\vec{r}_j)^2$$

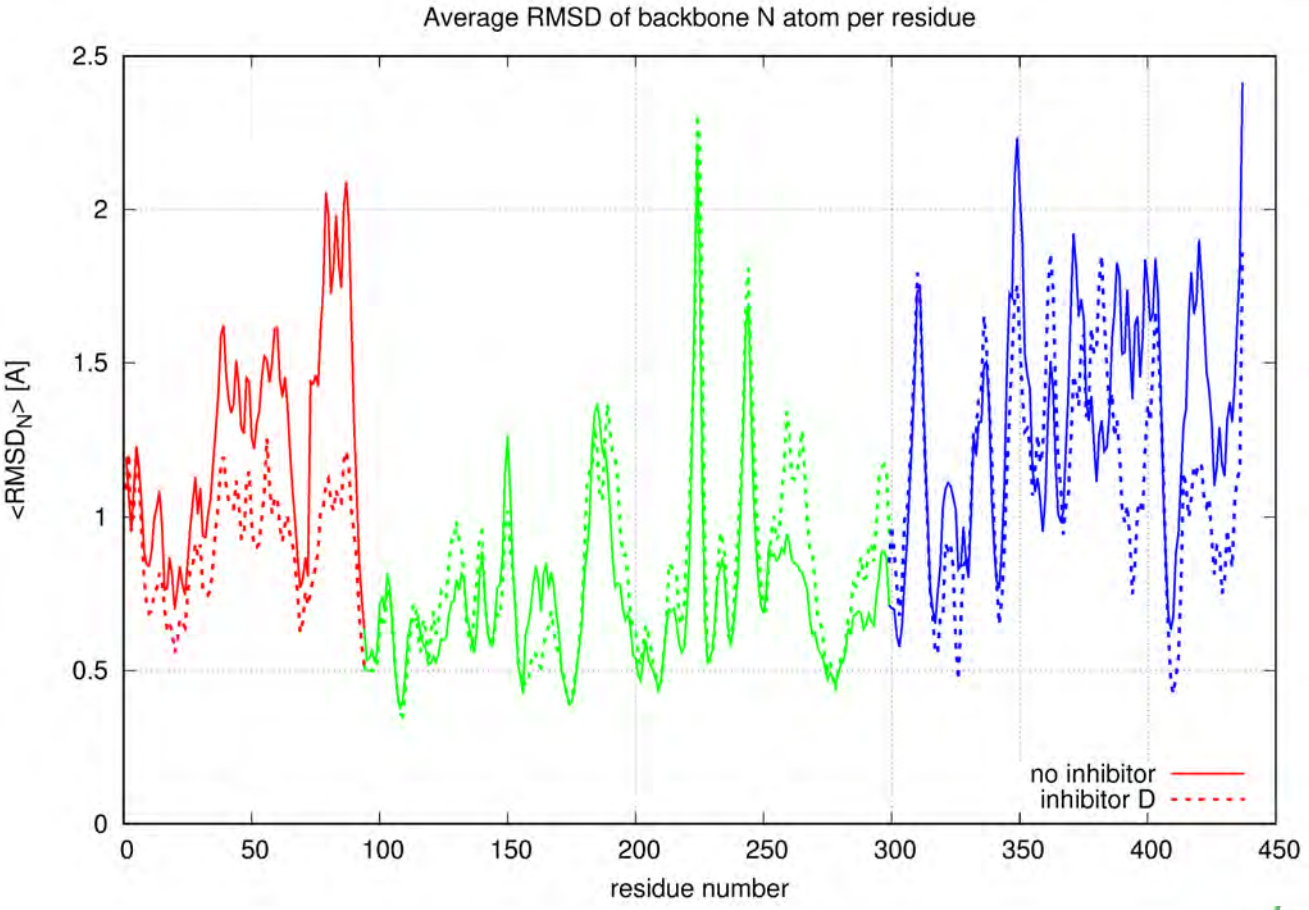
Example: development of the antibacterial agents – inhibitor of the **MurD protein**





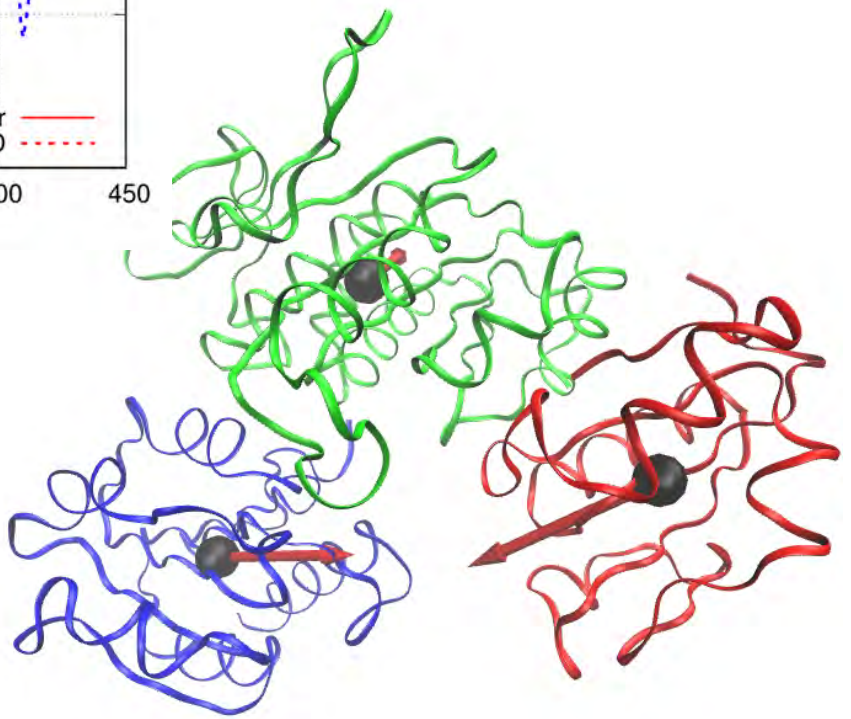
System with ~36.000 atoms

Solvent: wat + DMSO: ~30.000 atoms



Analysis of the 60ns all-atom MD

(~1 month & 20 HPC cores)



Internal motion (vibrations) analyzed with the Principal Component Analysis (PCA)

Shown is lowest (slowest) vibrational mode of inter-domain motion

ω(cmplx with inhibitor) = 8.2 10^{10} s^{-1} > ω(no inhibitor) = 5.3 10^{10} s^{-1}

General properties of the classical force fields (FF)

Good points

- Fast and easy to use
- Visualization of microscopic behavior

Weaknesses

- Quality of the FF difficult to asses
- Chemical reactions difficult to model (bonds are fixed)
- Orbital interactions (polarizability) often not included

Ab-initio MD simulations
(normally inevitable in bond-beaking/creating processes)

FF simulations might not be appropriate as they omit polarization effects!!

General simulation protocol remains the same as is used in the classical MD simulations.

Instead from the empirical potential function, interatomic forces are calculated from the DFT derived potential energy function (Hellmann-Feynman forces).

Car-Parrinello method.

Calculations based on direct minimization of the Kohn-Sham equations at each time step are now dominating. This approach is implemented in the program package VASP.

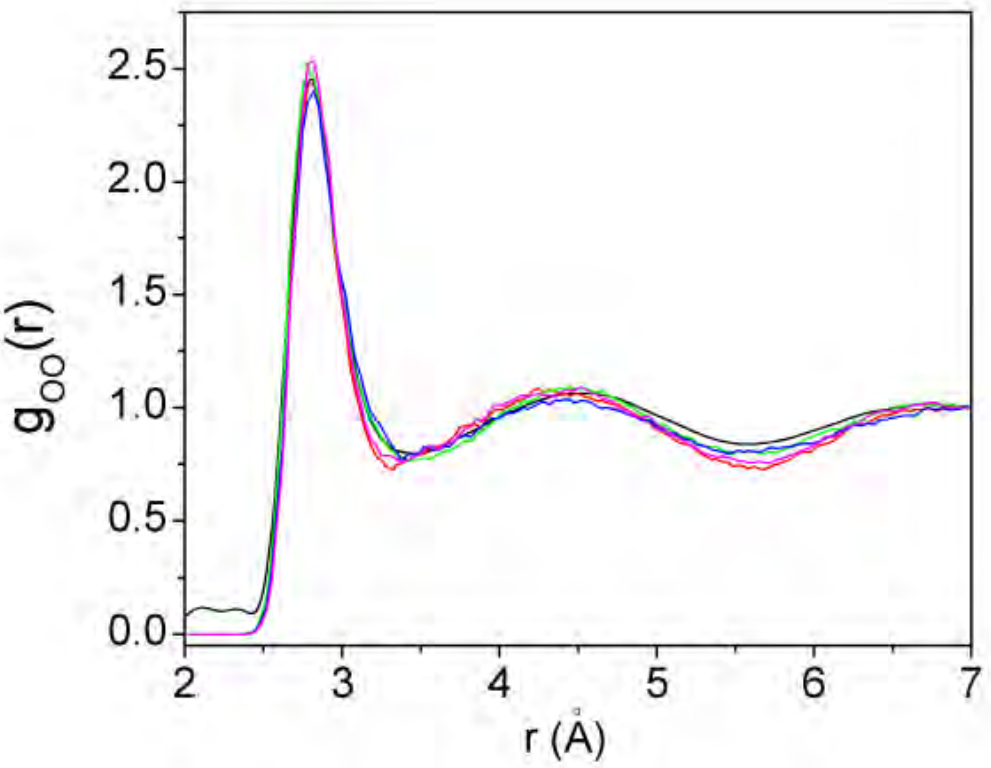
Choice of the proper DFT functional

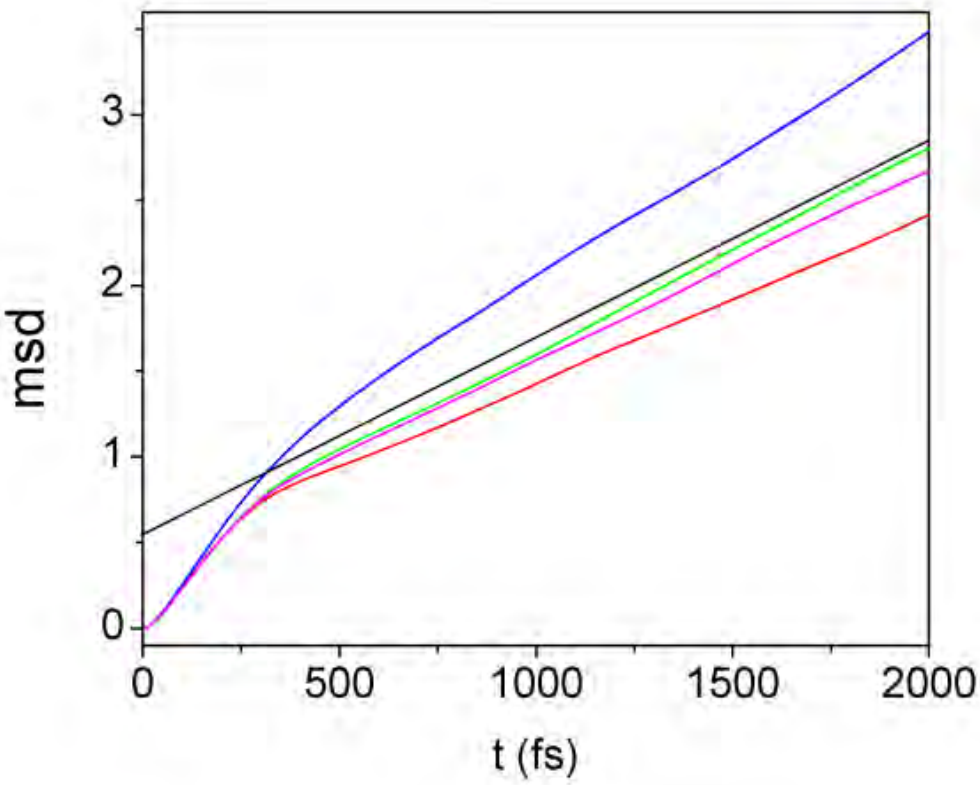
AIMD (ab-initio MD) choice of the appropriate functional (protocol) applying **vdW correction**

VASP program: plane-wave implementation of the density functional theory (DFT) in combination with the projector augmented wave pseudo-potential (PAW).

The choice of various DFT functionals possible

Testing the choice against reproducibility of the g(r) and diffusion constant





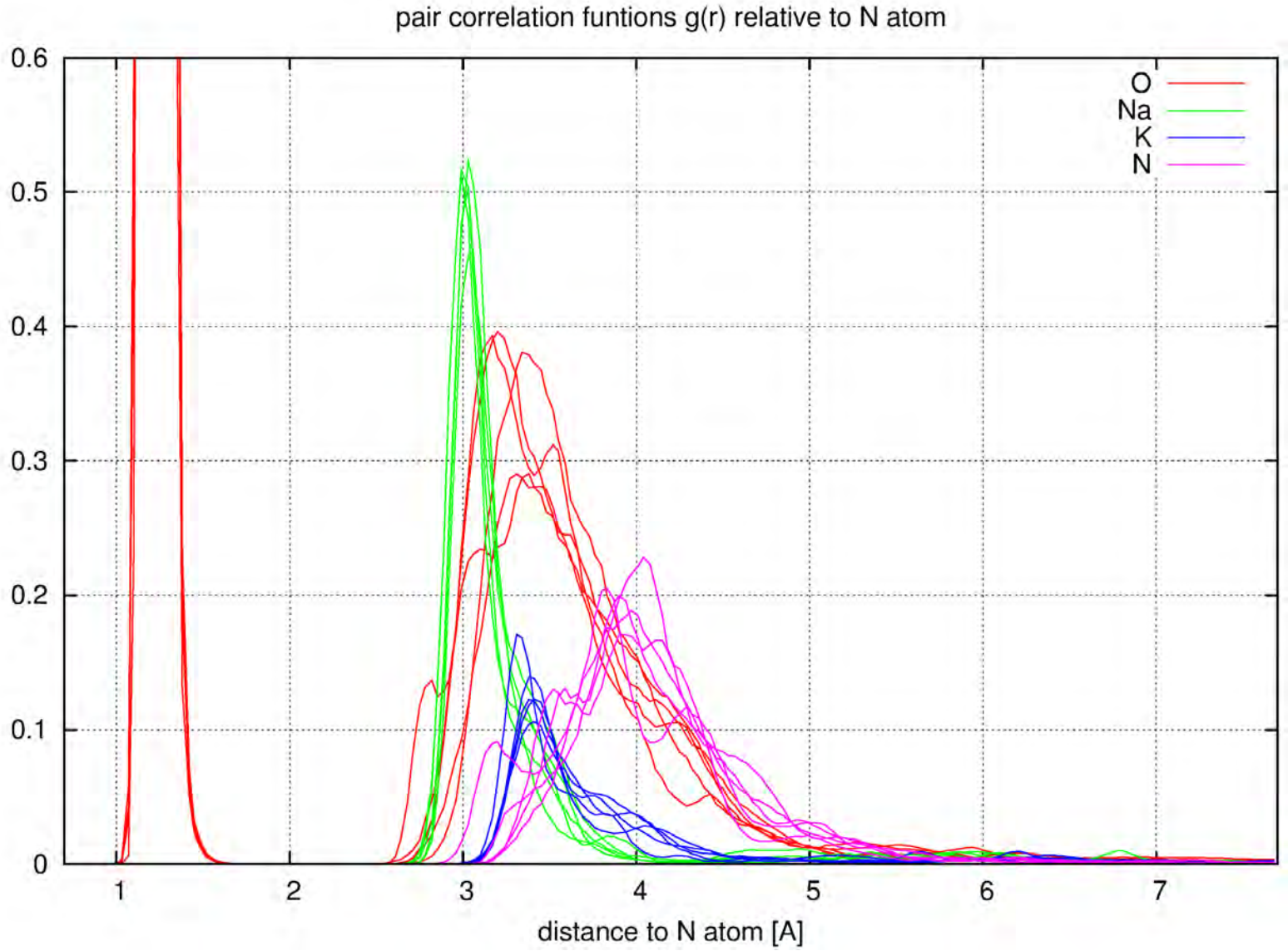
Black curves: experimental g(r): Skinner, black line with the slope of the experimental D

red: optPBE-DRSLL, **green: revPBE-D3**, blue: RPBE-D3, magenta: PBE-DRSLL

revPBE functional + Grimme D3 correction (for dispersion forces)

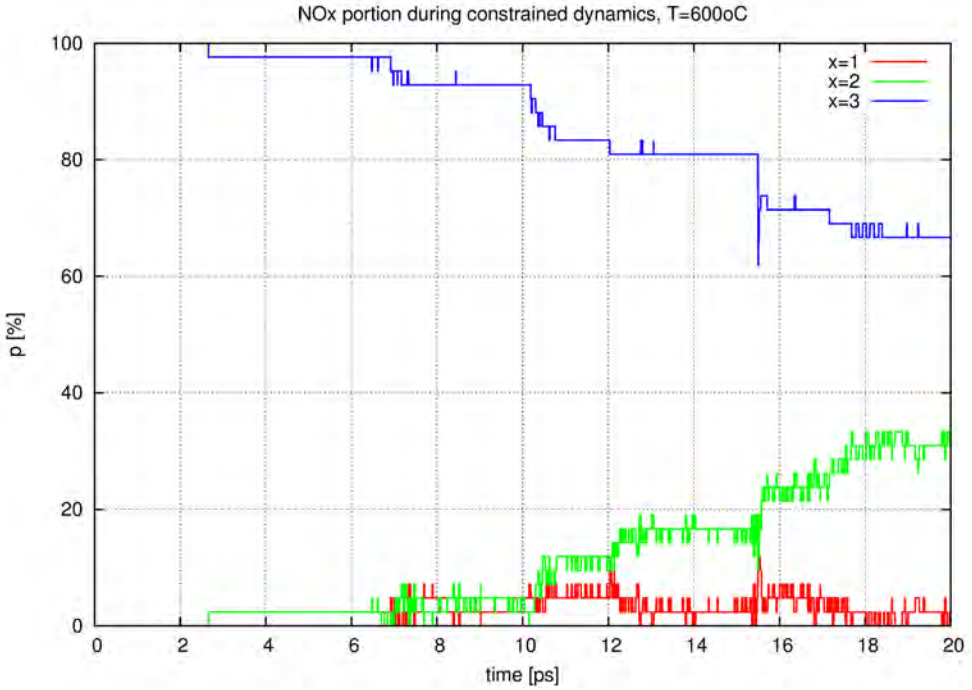
Molten (solar) salts simulations (KNO_3:$NaNO_3$)

Evaluation of atomic pair correlation functions $g_{NO}(r)$, $g_{NNa}(r)$, $g_{NK}(r)$), $g_{NN}(r)$, for systems at different T. Analysis performed on the 10ps trajectories followed 10ps of equilibration.

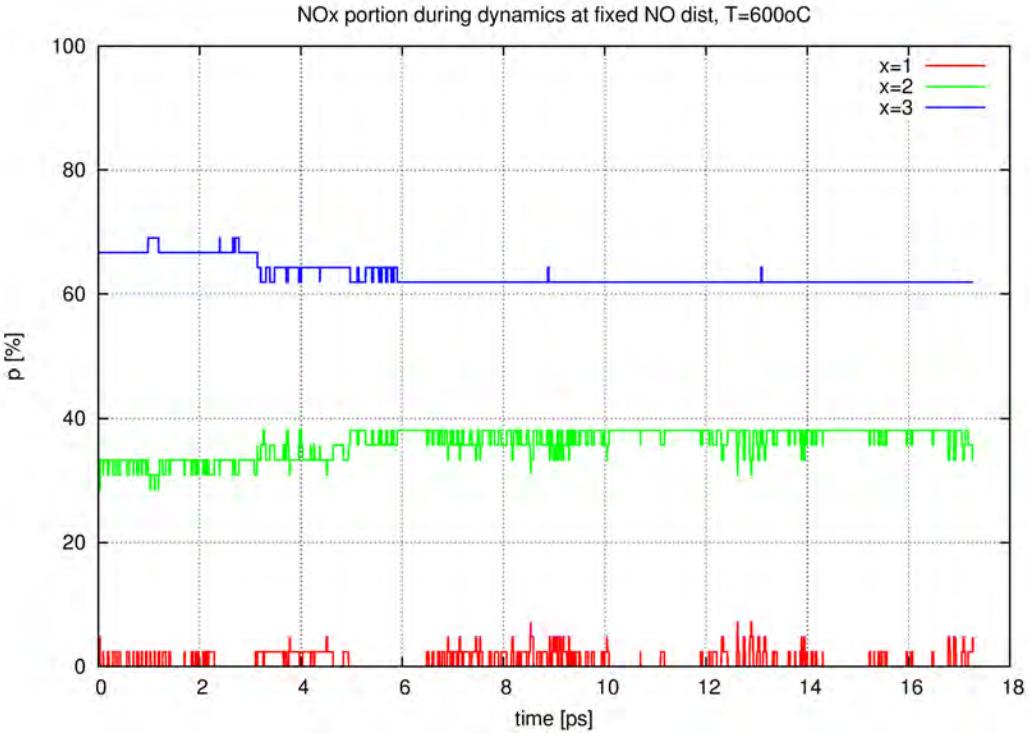


Degradation based on presence of the O-

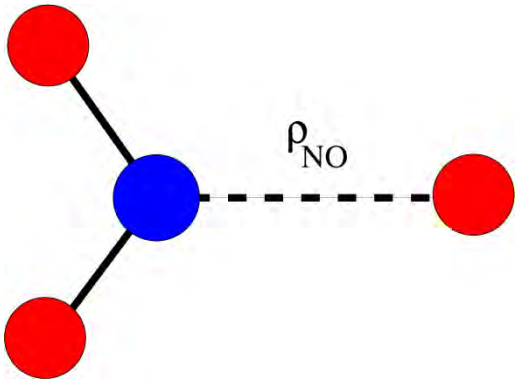
- Simualtions at T=600C

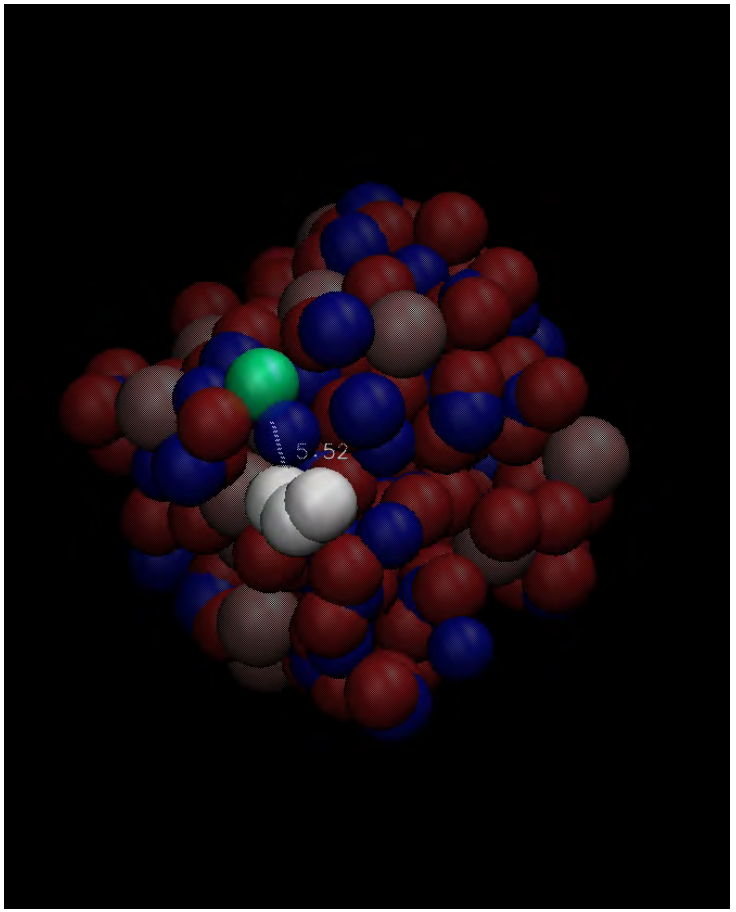


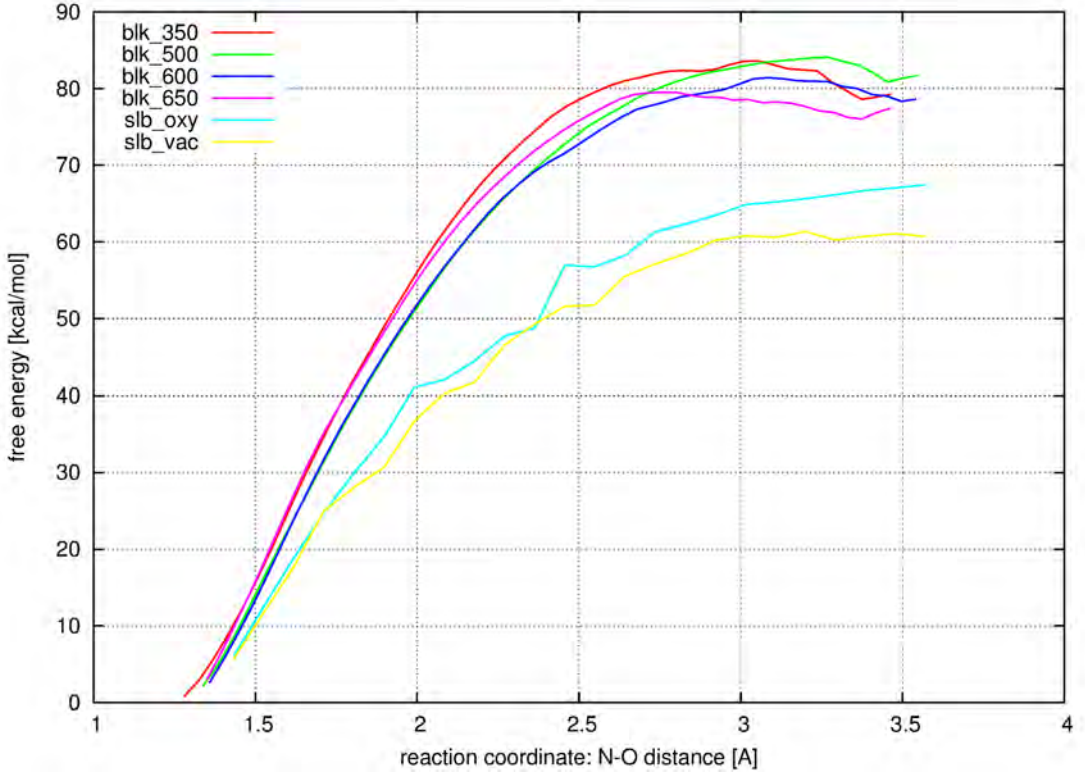
NO_x degradation



Analysis of the presence of NO (red), NO2 (green) and NO3 (blue) in two subsequent simulations.







Thermodynamic integration
using **constrined dynamics** for the reactioncoordinate(N:Odistance):

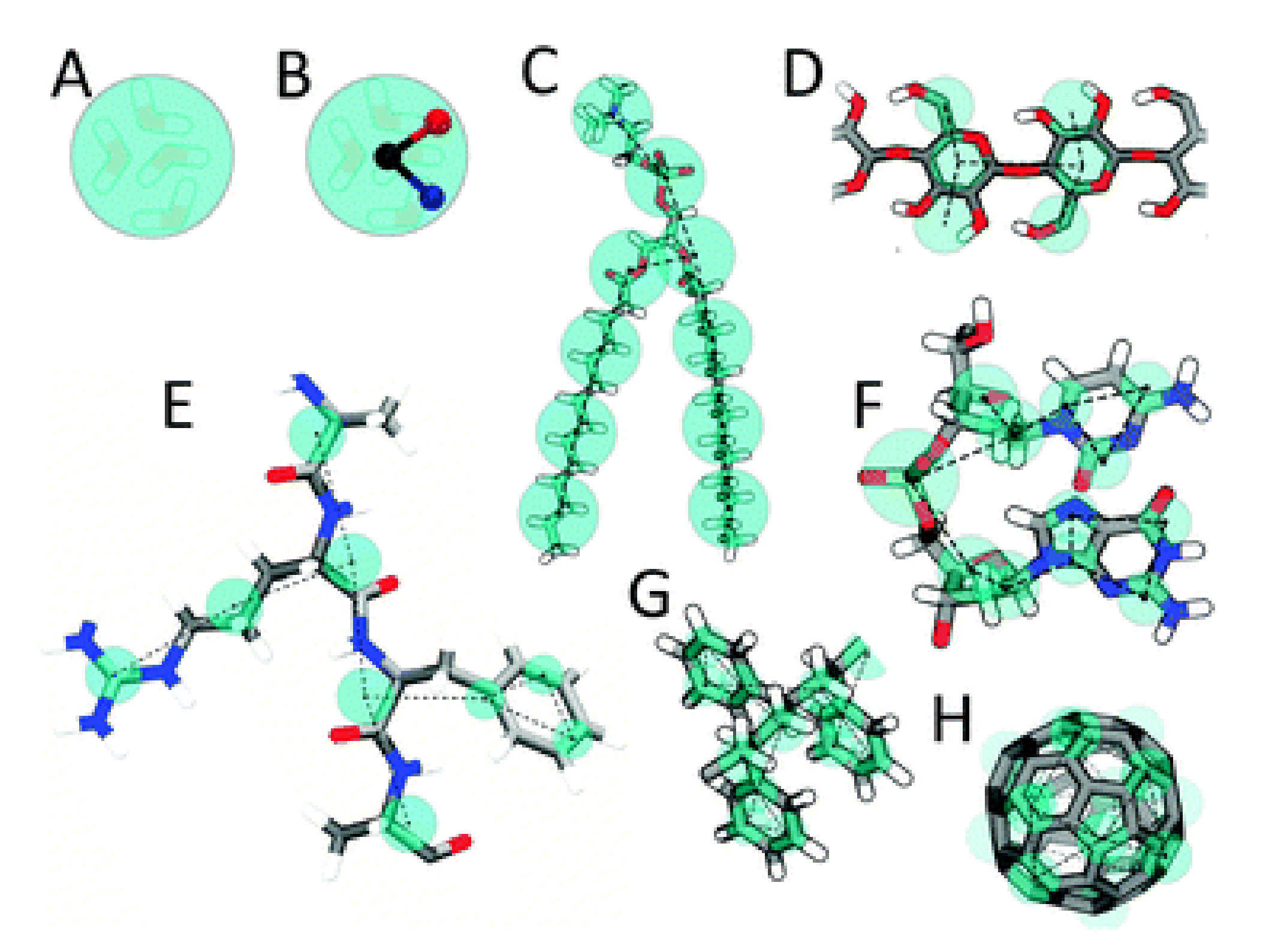
$$\Delta A = \int_{\xi_i}^{\xi_f} \left(\frac{\partial A}{\partial \xi} \right) d\xi$$

COARSE GRAINING: reducing degrees of freedom

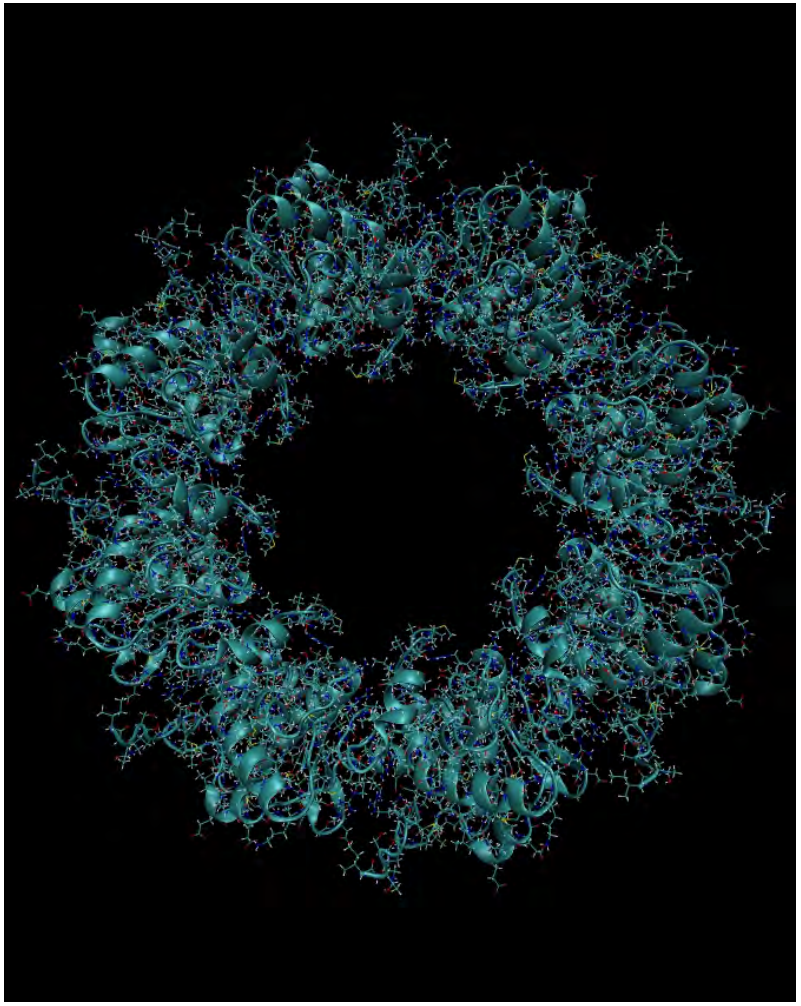
4-to-1 mapping: four heavy atoms and associated hydrogens are represented by a single Interaction center

only four main types of interaction sites: polar, non-polar, apolar and charged

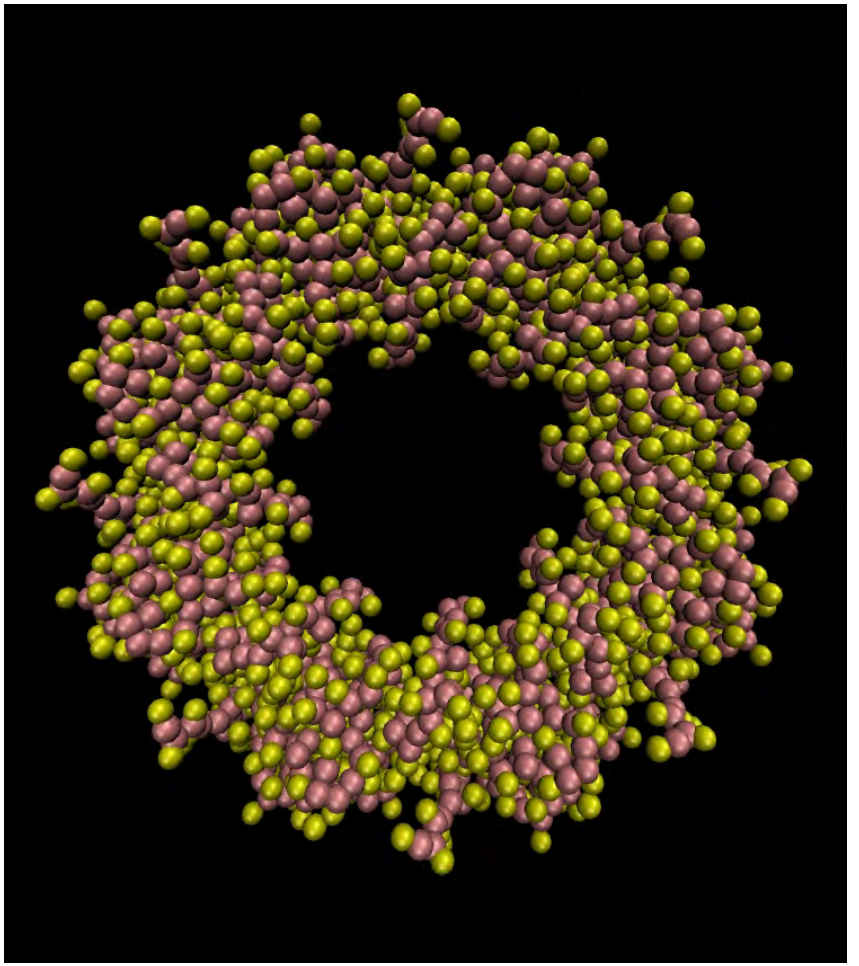
a number of subtypes allow for an accurate representation of the chemical nature of the underlying atomistic structure


A
B
C
D
E
F
G
H

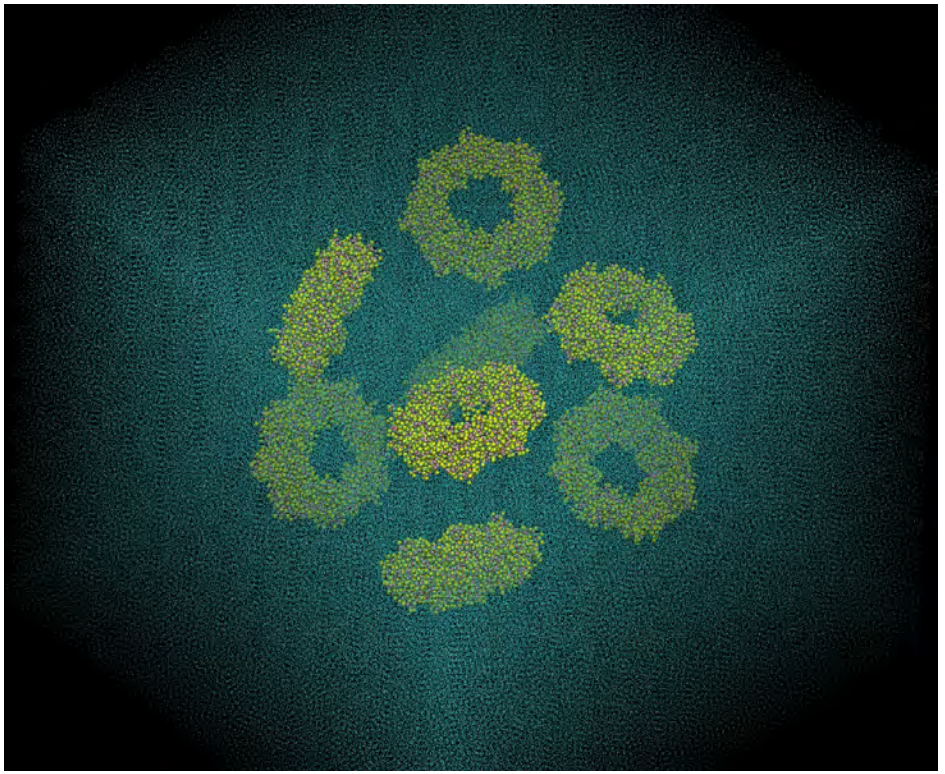
Simulating packing of virius-like-particles (VLPs)



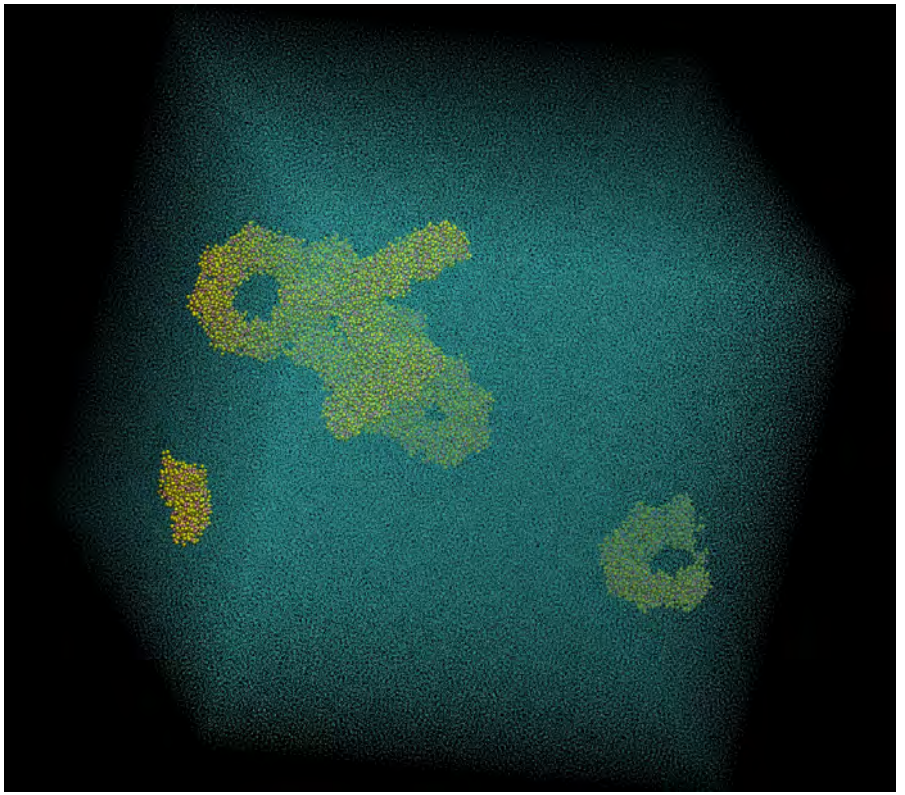
all-atom representation
~23.000 atoms



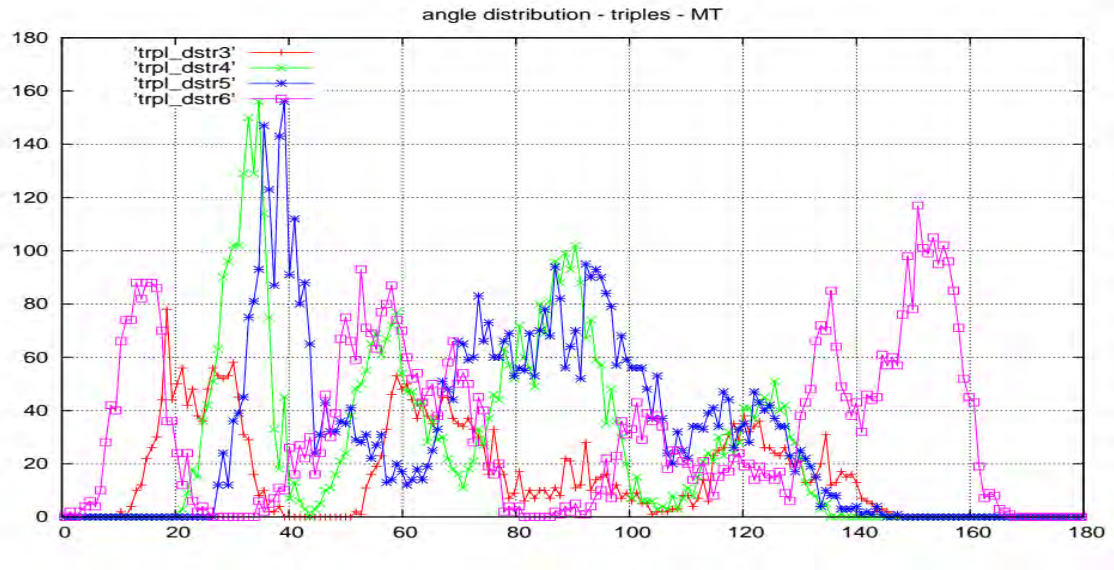
coarse-grained representation
~3.000 beads



8 rings at the beginning of simulation



Snap-shot after 5mic. Sec of sim



Conclusions > some caveats about computer simulations:

- computer simulations are extremely powerful tool in chemistry, but

- (almost always) involve only a limited degrees of freedom → finite size effects

- involve a limited time scale → insufficient sampling to extract reliable thermodynamic quantities

- interactions between particles are not exact and often approximated (long range) → systematic lack the description accuracy

- crucial point of any simulation is therefore a critical assessment of its results, optimally against the ability to reproduce experimental data

Neli Sedej, Ph.D. student at FMF, UL

Marko Sever, Ph. D. student ant FFA, UL