Computer simulation and neutron scattering

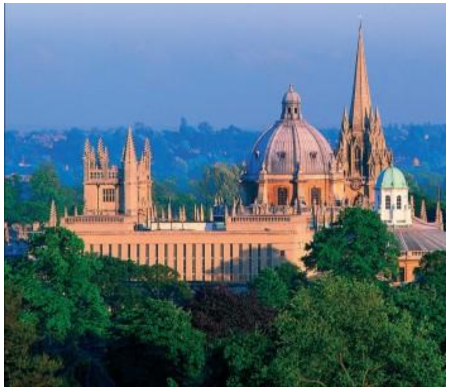
Miguel A. Gonzalez

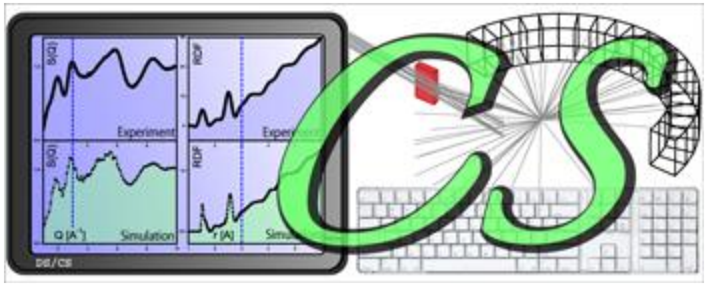
Computing for Science

Institut Laue Langevin, Grenoble, France

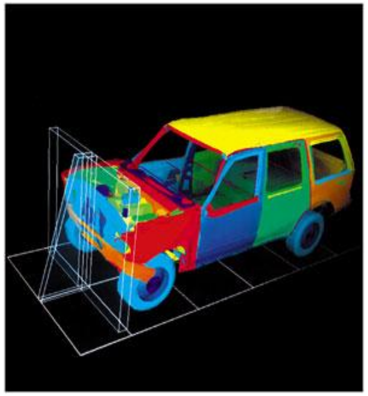
13th Oxford School on Neutron Scattering

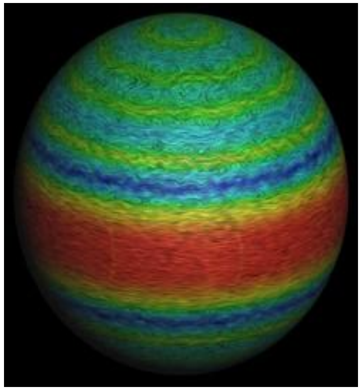
2-13 September 2013





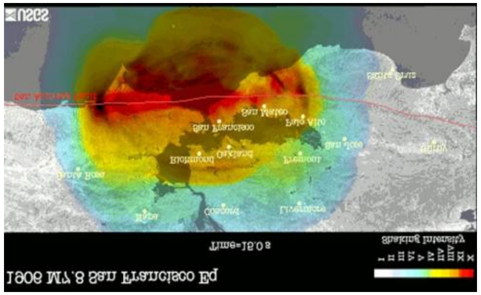
Why computer modeling?





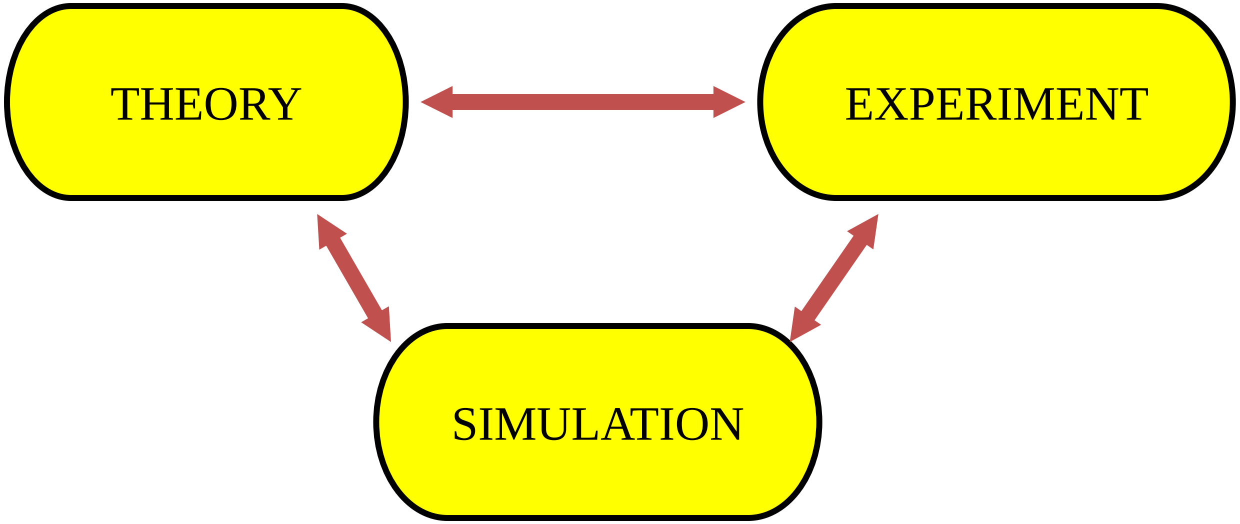
- To replace experiments when they are impossible or dangerous
    - Astrophysics, nuclear accidents, earthquakes, ...
    - High temperatures and pressures
    - No stable phases (e.g. water in "no man's land")

- If a good **model** is available, they are a cheap and easy way of obtaining reliable data
    - Rapid development of hardware and software
    - Minimize and optimize expensive experiments



- They can provide unique information
    - Behavior of particular atoms
    - Compute special correlation functions
    - Test theories (normally based on ideal models)

Why computer modeling?



THEORY

EXPERIMENT

SIMULATION

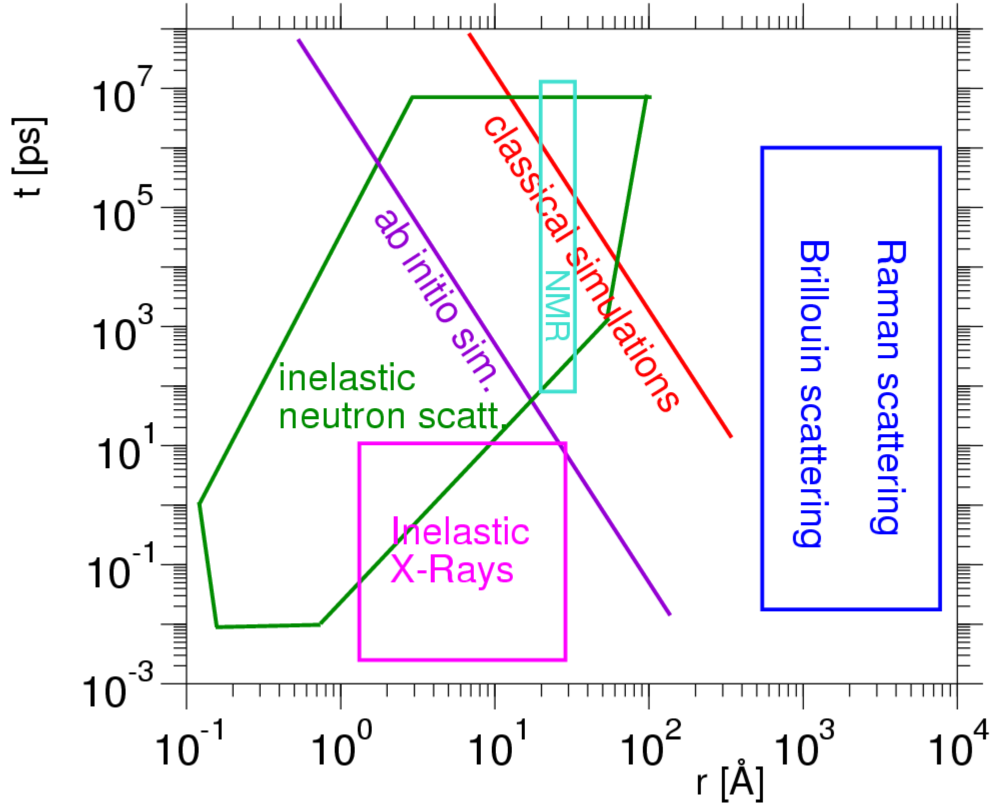
But don't forget that we do not simulate the real system, but a **model**: $N_{atoms} << 10^{23}$, simulation time << observation times in experiment, approximate potential, etc.

We use simulations to improve the experiments to perform (e.g. multiple scattering, points to measure) and to understand better the experimental results!

Simulations and neutrons

- Neutrons see nuclei
  - Van Hove correlation functions: $S(Q,\omega) \leftrightarrow F(Q,t)$
  - Much more direct link between simulations and neutron scattering that with other techniques (specially for classical MD)

- Same (Q,ω) range
  - time: fs to ns
    (up to μs in some cases;
    < 1 ns if ab initio)
  - length: Å to several nm
    (few Å with ab initio DFT)



What kind of simulation?

It will depend on the physics to study and the computer means available.

- The model and the method of solution will depend on the purpose of the simulation: they should be **accurate** and **efficient**.
- Accurate means that the simulation will reliably predict the behavior of the real system.
- Efficient means feasible with the available technical means.

What kind of simulation?

Choose way of computing interatomic forces

Hartree-Fock methods

Semiempirical methods

Density Functional Theory

Tight-binding

Empirical force fields

Coarse graining

Brownian Dynamics

Langevin Dynamics

Dissipative Particle Dynamics

Choose way of sampling the phase space

Lattice Dynamics

$$D_{ij}(\mathbf{k}) = \frac{1}{\sqrt{m_i m_j}} \sum_{\ell} \Phi_{ij}(0\ell) \exp\left(i\mathbf{k} \cdot [\mathbf{r}_i(0) - \mathbf{r}_j(\ell)]\right)$$

Thermodynamics

Phonons

Monte Carlo

$$\pi_a(\Gamma \rightarrow \Gamma') = \min(1, \exp(-\beta \Delta H))$$

Thermodynamics

Structure

Molecular Dynamics

Levels of approximation

Compromise between computational cost, accuracy and generality

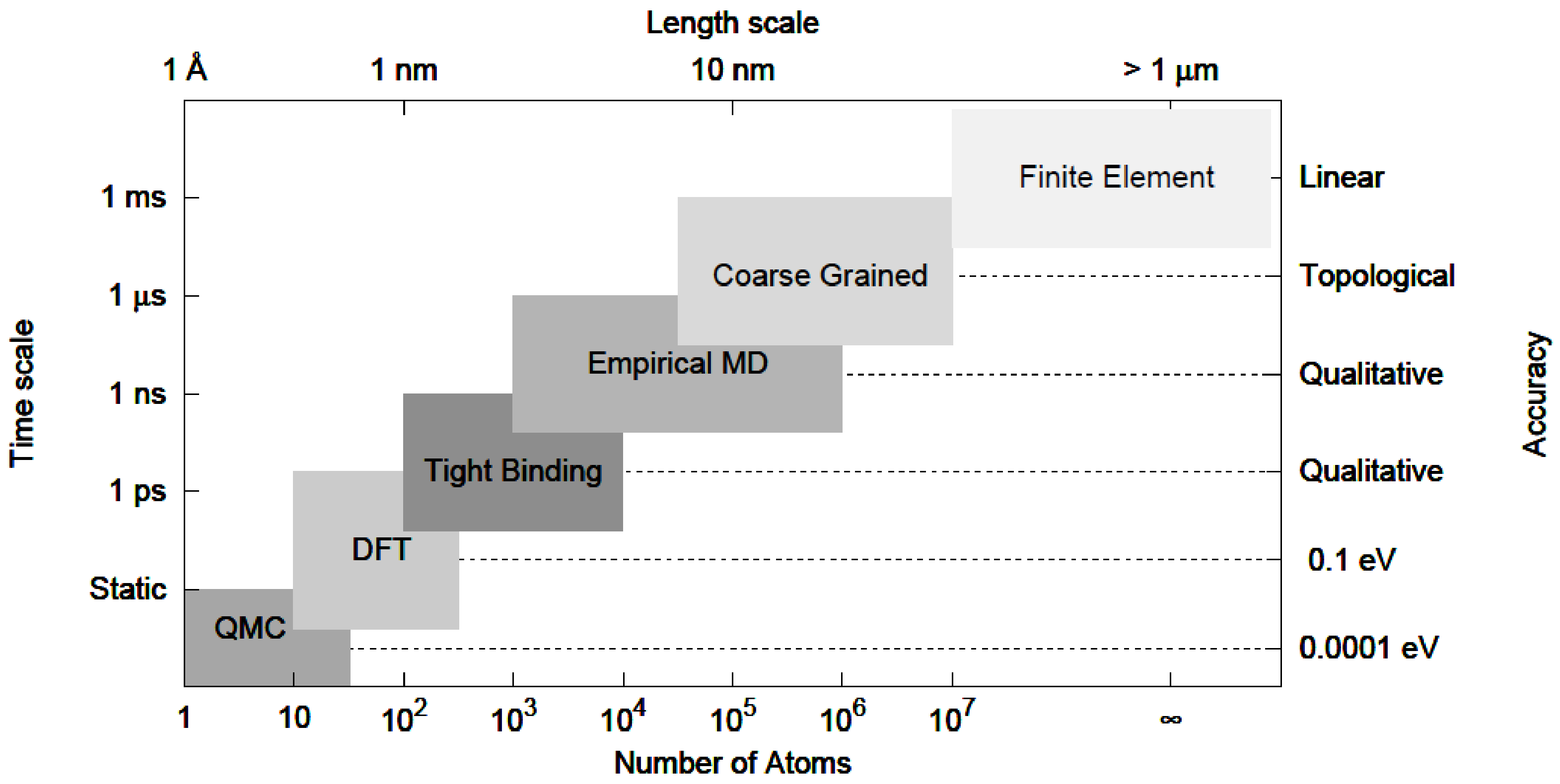


Figure 1. Schematic representation of the range of length- and time-scales accessible to a variety of modelling methods, from quantum Monte Carlo (QMC) for very accurate, very expensive static calculations through to approximate methods such as finite-element modelling.

From Kermode et al., in *Multiscale Simulation Methods in Molecular Sciences,* J. Grotendorst, N. Attig, S. Blügel, D. Marx (Eds.), NIC Series, Vol. 42, pp. 215-228 (2009).

Electronic structure

Solve Schrödinger equation: $\hat{H}\Psi = E\Psi$

Generally under Born-Oppenheimer approximation (and non relativistic)

$$\hat{H}\Psi = \left[-\frac{\hbar^2}{2m}\sum_{i}^{N}\nabla_i^2 + \sum\sum_{i<j}^{N} U(\underline{r}_i, \underline{r}_j) + \sum_{i}^{N} v_{ext}(\underline{r}_i) \right] \Psi = E\Psi,$$

$$U(\underline{r}, \underline{r}') = e^2 |\underline{r} - \underline{r}'|^{-1}$$

v_{ext} = external potential containing the interaction of the e^- with the fixed nuclei + constant term from nuclear-nuclear interaction

- Hartree-Fock and post Hartree-Fock methods (CI and QMC)
- Limited to small molecules (geometries and vibrations) and crystals with small unit cells.
- Semi-empirical (MINDO, AM1, …): Based on HF formalism, but with many approximations and some parameters from experiment.

Electronic structure: DFT

- Based on Hohenberg-Kohn and Kohn-Sham theorems.
- The electron density determines uniquely the properties of the ground state.
- It is possible to define an energy functional E[n(r)] and the ground state electron density minimizes this functional.
- The many-body problem of finding the minimum of $\langle\Psi|\hat{H}|\Psi\rangle$ for many-electron trial wavefunctions is reduced to the simpler one of finding the minimum of E[n(r)] for trial densities n(r) which depend on only 3 space variables.

$$E[n(\underline{r})] \leq \left\langle \Psi \middle| \hat{T} + \hat{U} + v_{ext} \middle| \Psi \right\rangle \text{ and } E_0 = \min E[n(\underline{r})]$$

- The energy functional is defined as: $E[n(\underline{r})] = F[n(\underline{r})] + \int n(\underline{r}) v_{ext}(\underline{r}) d\underline{r}$
- F[n(r)] is unknown, but can be expressed as a sum of known terms and into an unknown (hopefully much smaller) term which must be approximated:

$$E[n(\underline{r})] = T[n(\underline{r})] + \int n(\underline{r}) v_{ext}(\underline{r}) d\underline{r} + \frac{e^2}{2} \iint \frac{n(\underline{r}) n(\underline{r}')}{|\underline{r} - \underline{r}'|} d\underline{r} d\underline{r}' + E_{xc}[n(\underline{r})]$$

- E_{xc}[n(r)] is the exchange-correlation energy functional which is not known and must be approximated in a reasonable way.

Electronic structure: DFT

- Need to find useful, approximate functionals for $E_{xc}[n]$

- Common functionals:

  - LDA (Local Density Approximation):
    - Uses only n(r) at a point
    - Assumes the functional is the same as in the homogeneous electron gas

  - GGA (Generalized Gradient Approximation)
    - Uses n(r) and $|\nabla n(r)|$
    - Generally more accurate, corrects overbinding of LDA
    - PBE, BLYP, AM05

  - Hybrid
    - Add some fraction of HF
    - B3LYP, PBE0

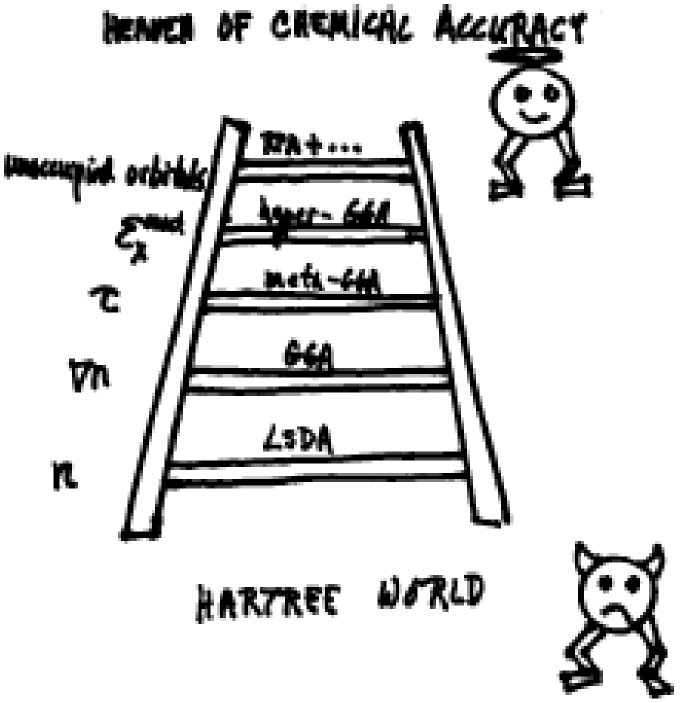


Figure 2. The Jacob's ladder of density functional approximations to the exchange-correlation energy adds local ingredients successively, leading up in five steps from the Hartree world ($E_{xc} = 0$) of weak or no chemical bonding to the heaven of chemical accuracy (with errors in energy differences of order 1 kcal/mol=0.0434 eV).

From Perdew, *J. Chem. Theory Comput.* **5**, 902-908 (2009).

Electronic structure: DFT

Which functional should I use?

D. Rappoport, N. R. M. Crawford, F. Furche, and K. Burke.

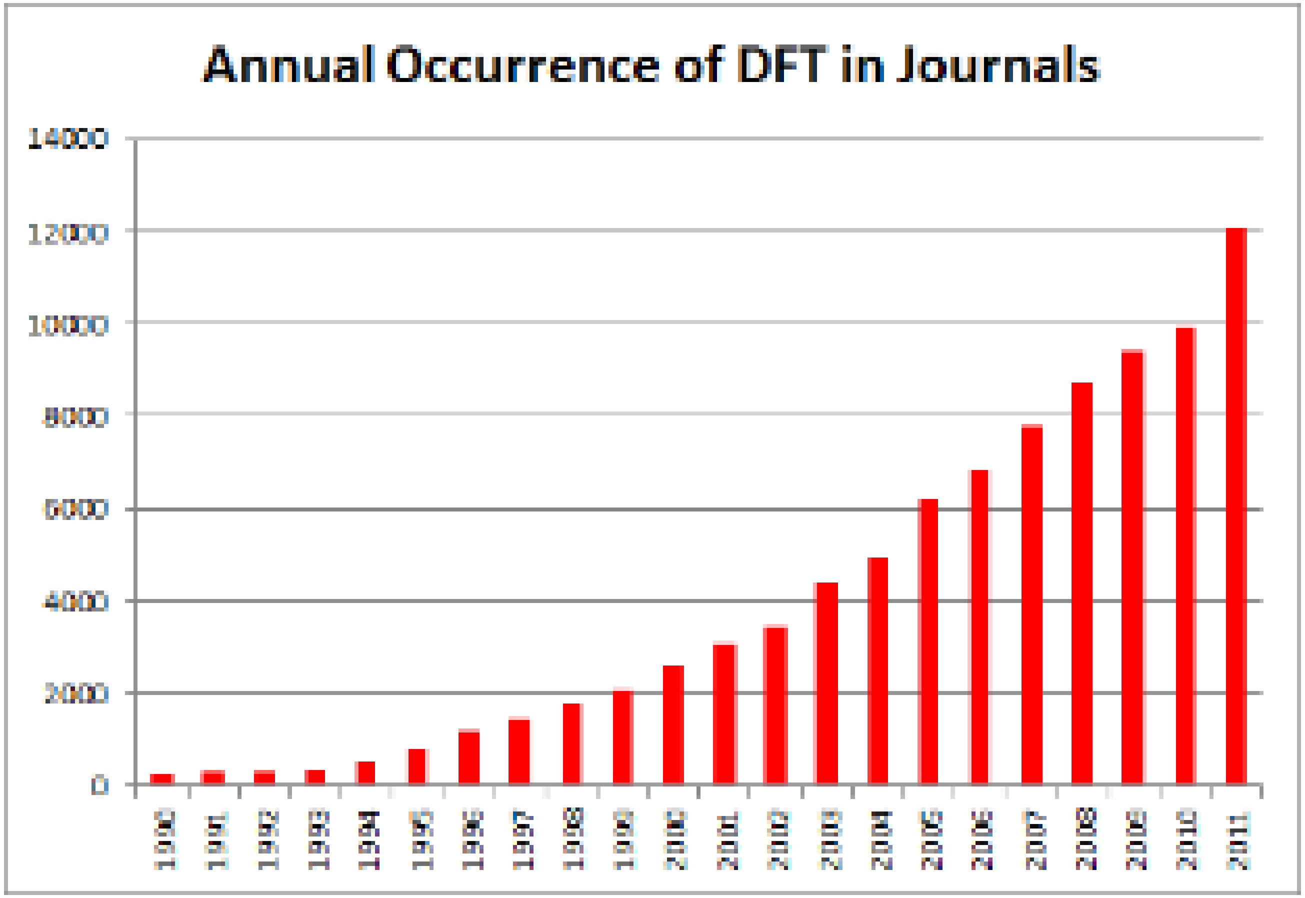
In "Computational Inorganic and Bioinorganic Chemistry", E. I. Solomon, R. A. Scott, and R. B. King (eds), Wiley-Blackwell, 2009.

Available in http ://dft.uci.edu/pubs/RCFB08.pdf.

- *All functionals used in practice are approximations.*
- *No presently existing functional is highly accurate for all properties of interest.*
- *Good non-empirical functionals are widely applicable, but good empirical functionals are often more accurate (at least for properties and systems they've been designed for).*

"Clearly, there is no single answer to the title question. At any given time, and for any given property and system, there is at most a "best" answer. Experience and benchmarking are always needed to find that best answer."

Electronic structure: DFT

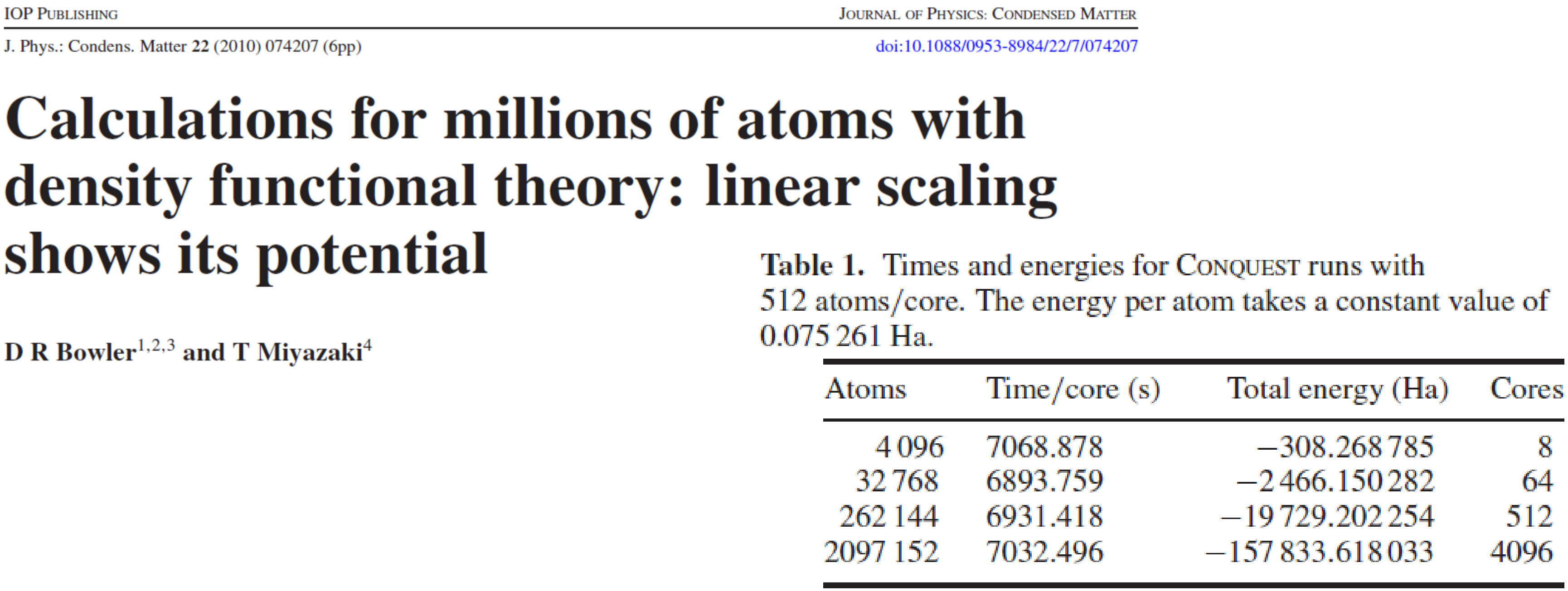


- Today DFT is the reference method in materials simulation
- Tipically handles systems containing ~100-1000 atoms, times ~100 ps.

Electronic structure: DFT

Problems:

- Incomplete treatment of dispersion forces → Hybrid functionals
- Strongly correlated systems → LDA+U, DMFT, …
- Band gaps in semiconductors underestimated → Hybrid functionals
- Limited system size → linear scaling DFT



IOP PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **22** (2010) 074207 (6pp) doi:10.1088/0953-8984/22/7/074207

Calculations for millions of atoms with density functional theory: linear scaling shows its potential

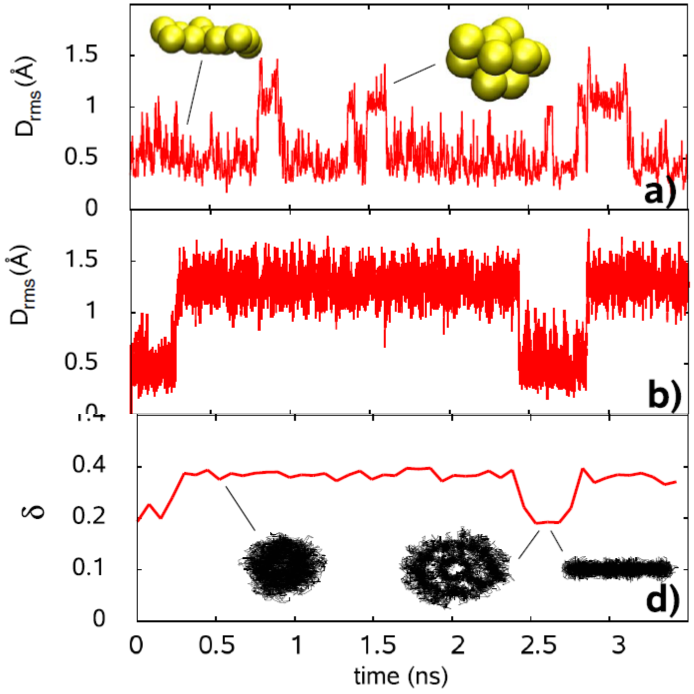
D R Bowler[1,2,3] and T Miyazaki[4]

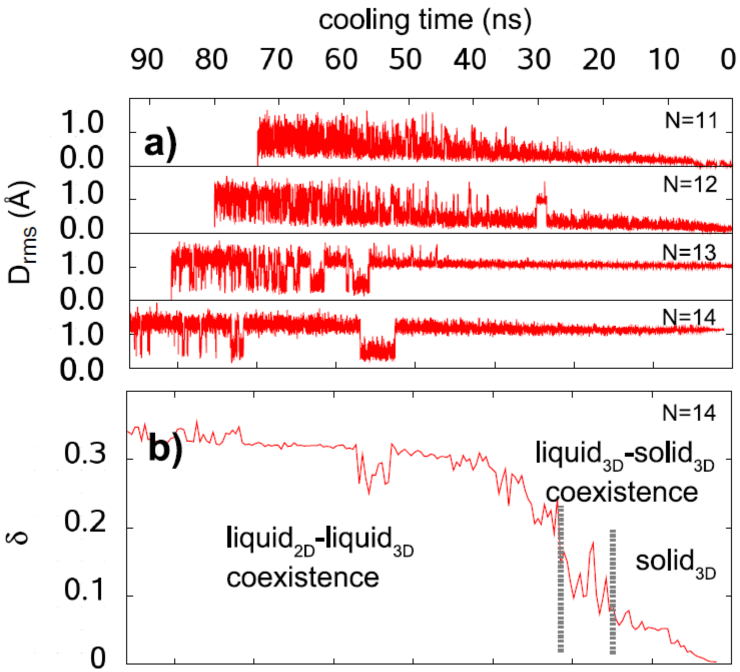
Table 1. Times and energies for CONQUEST runs with 512 atoms/core. The energy per atom takes a constant value of 0.075 261 Ha.

Atoms	Time/core (s)	Total energy (Ha)	Cores
4 096	7068.878	−308.268 785	8
32 768	6893.759	−2 466.150 282	64
262 144	6931.418	−19 729.202 254	512
2097 152	7032.496	−157 833.618 033	4096

Electronic structure: Tight Binding

- Assumes electrons are tightly bound to the atom to which they belong and have limited interaction with surrounding atoms.
- The wave function of the electron will be similar to the atomic orbital of the free atom.
- Many of the matrix elements of the tight binding Hamiltonian are parameterized from experiment or using DFT (DFTB).
- 3 orders of magnitude faster than DFT, but lost of accuracy and transferability.

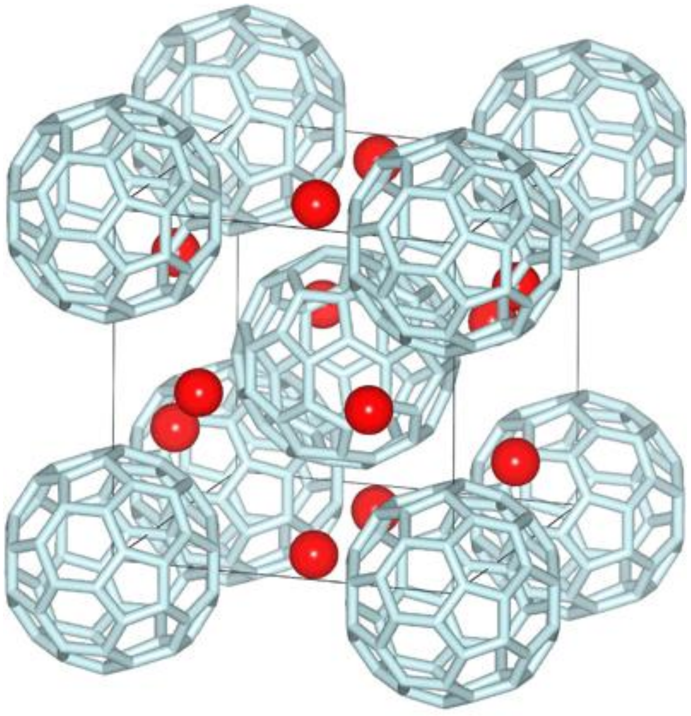


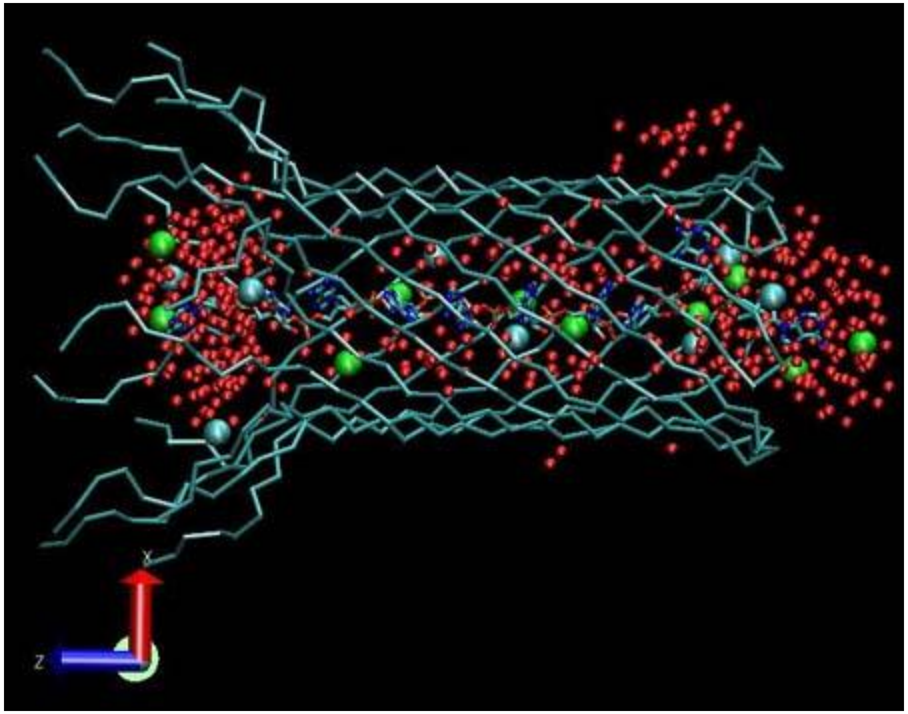


Koskinen, PRL (2007)

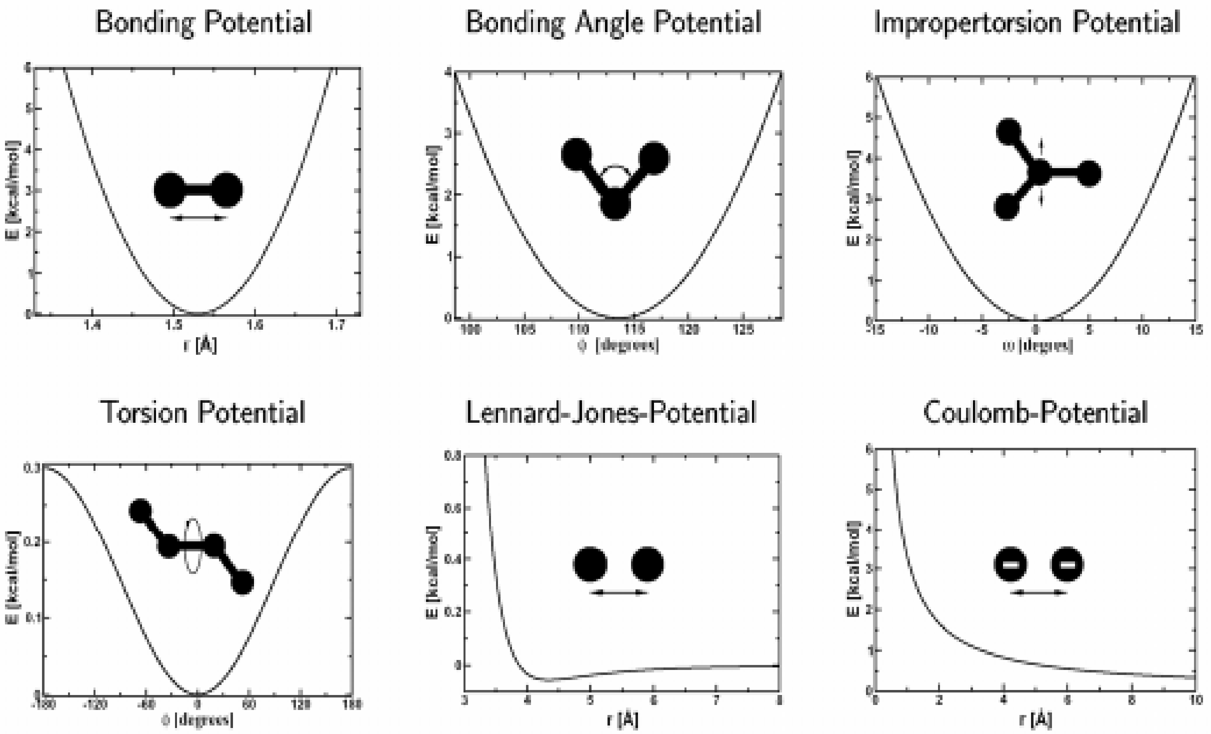
Molecular Mechanics

- No electrons
- Molecules represented as ‘balls’ + ‘springs’
- Empirical potential to account for intra- and intermolecular interactions
- ~5-6 orders of magnitude faster than DFT → ~10^4 atoms, ~1-10 ns





Empirical potentials



$$
\begin{aligned}
E \quad = \quad & \sum_b k_b (r - r_\mathrm{b})^2 + \sum_\theta k_\theta (\theta - \theta_\mathrm{O})^2 + \sum_\omega k_\omega (\omega - \omega_\mathrm{O})^2 \\
+ \quad & \sum_\phi k_\phi (1 - \cos(n\phi - \delta)) + \sum_{i<j} 4\ \epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r} \right)^{12} - \left(\frac{\sigma_{ij}}{r} \right)^6 \right] + \sum_{i<j} \frac{q_i q_j}{r}
\end{aligned}
$$

Limits of force fields

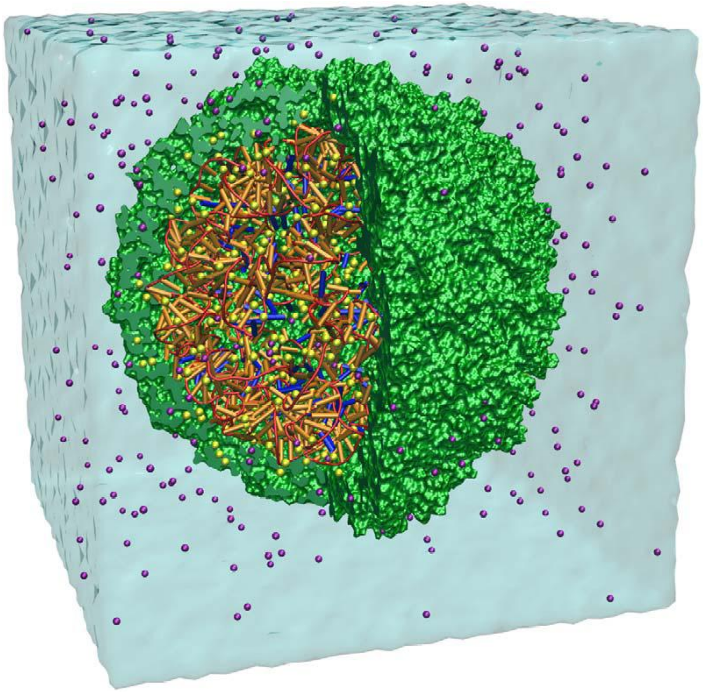
- **Intrinsic limitations:**
  - No information about the electronic structure.
  - No chemistry → Unable to handle reactions (bond-breaking/forming, electron excitation, charge transfer, etc.)
- **Limited prediction power:**
  - The accuracy depends on the parameterization.
  - It can only be used for systems having the functional groups that were included in parameterization.
  - Limited transferability: Be careful when applying a FF under conditions that are very different from the conditions used in its parameterization, e.g. P, T.

Needs experimental validation!

And possibilities …

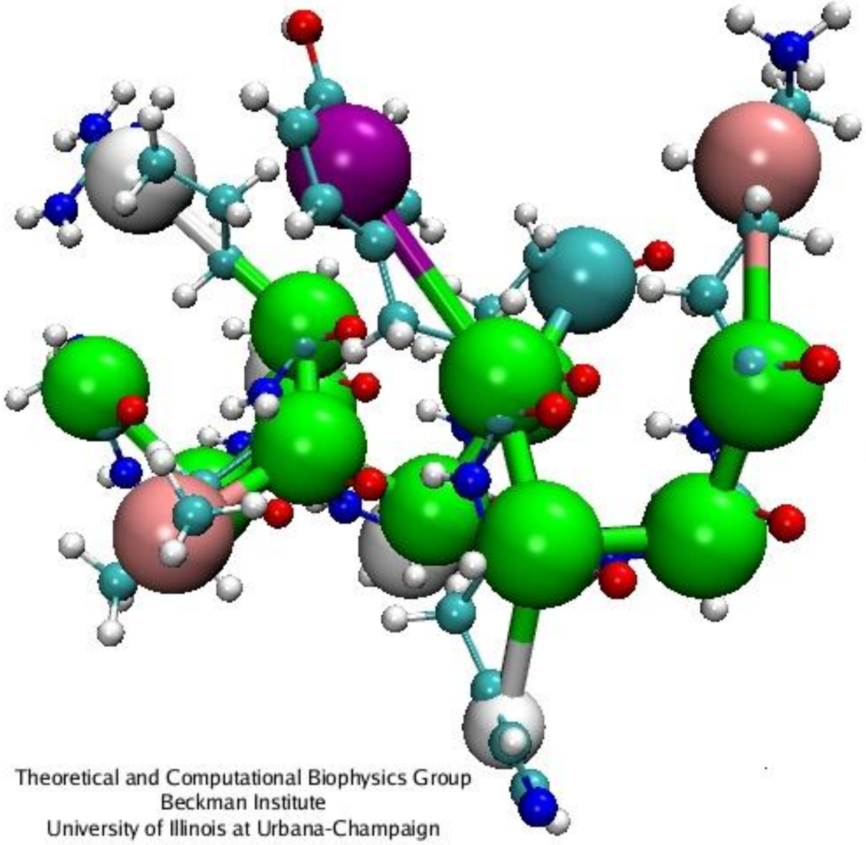
- **Analysis of energy contributions can be done at the level of individual interactions or classes of interactions.**
- **Possible to modify the energy expression to bias the calculation.**
- **Allow to handle large systems and simulate relatively long times (several orders of magnitude faster – and therefore cheaper – than quantum-based calculations):**

  - Small cluster (16 processors) ~ 10^4 atoms @ 1 ns/day
  - BlueGene/L (131072 processors) ~ 320 billion atoms (a cubic piece of metal of side ~1 μm) @ 10 ps/day (Kadau, *Int. J. Modern Physics C* (2006))
  - Anton (specialized machine for MD simulations) ~ 10^4 atoms @ 10 μs/day (Klepeis, *Curr. Opin. Str. Biol.* (2009))
  - NCS (256 nodes) ~ 10^6 atoms (all-atom satellite tobacco mosaic virus, NAMD) @ 1ns/day (Freddolino, *Structure* (2006))
  - Perspectives: multimillion-atom (~100 nm scale) @ 30 ns/day on a Cray XT5 (10^5 cores) (Schulz, *J. Chem. Theory Comput.* (2009))

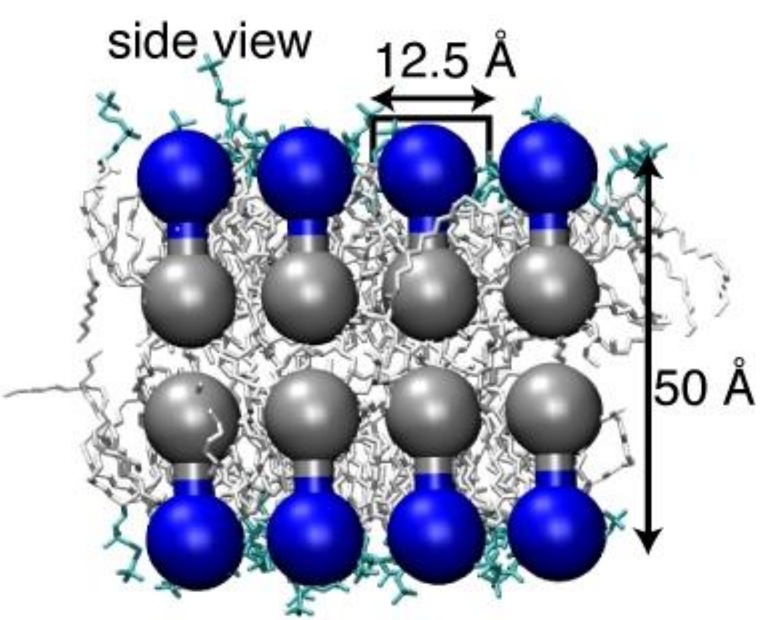


Coarse Graining

- Represent the system by a reduced number of degrees of freedom.
- E.g. protein-lipid CG model by Shih et al. (JPCB 2006): clusters of 10 atoms → single bead, 4 H_2Os → 1 ‘water’ bead, ion + solvation shell → 1 ‘ion’ bead, etc.
- Gain (several orders of magnitude) by reducing N and possibility to use larger Δt.
- Needs parameterization.



Theoretical and Computational Biophysics Group
Beckman Institute
University of Illinois at Urbana-Champaign



Langevin and Brownian Dynamics, DPD, …

- Remove ‘unimportant’ degrees of freedom, whose influence is replaced by a frictional force and a noise → Langevin dynamics:

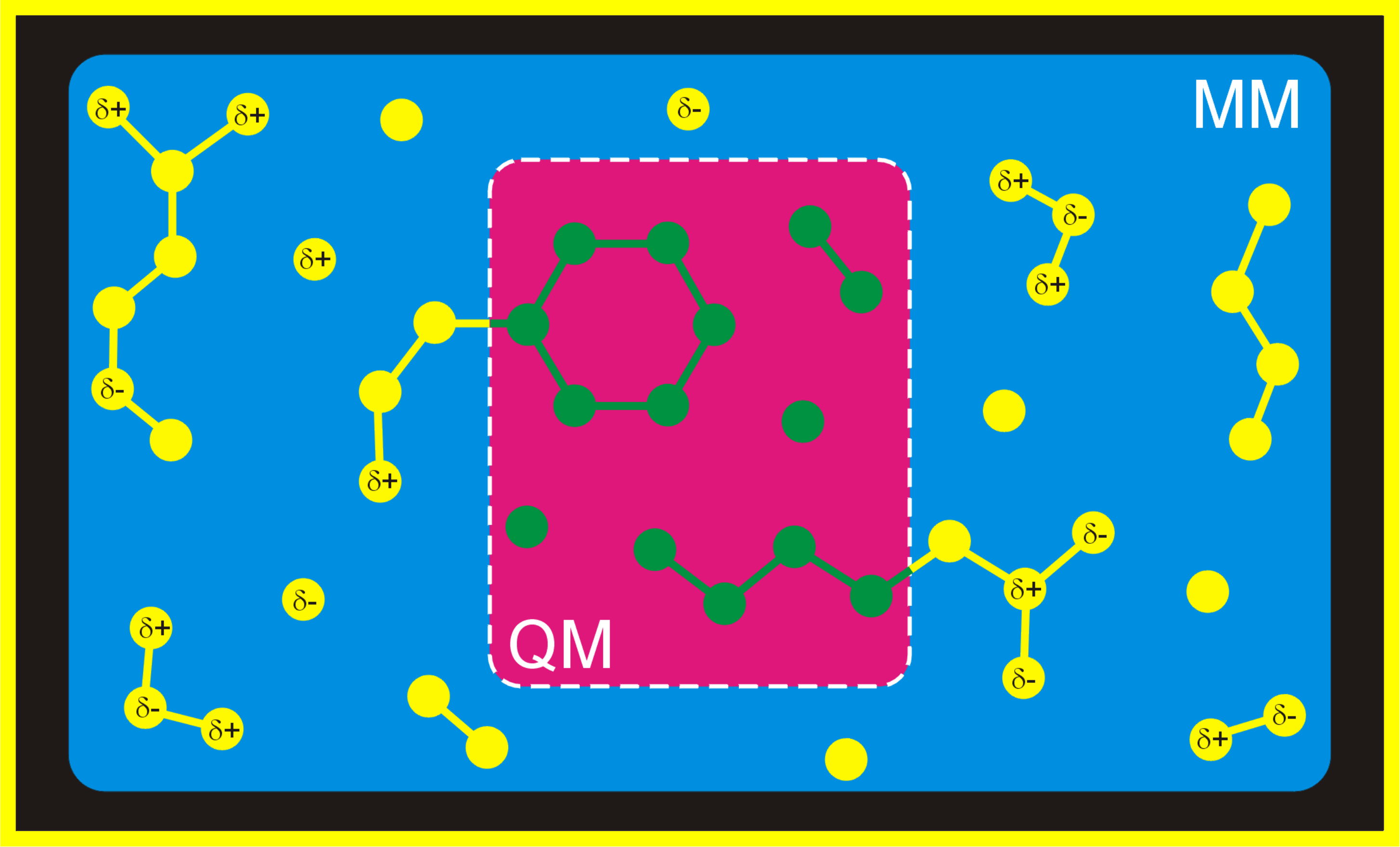
$$m_i \frac{d\vec{v}_i}{dt} = -\frac{\partial V^{mf}}{\partial \vec{r}_i} - m_i \sum_j \int_0^t \gamma_{ij}(\tau) v_j(t-\tau) d\tau + \vec{R}_i(t)$$

- If systematic force does not change much on the time scale of the VACF, the average acceleration becomes small and can be neglected→ Brownian dynamics:

$$0 \approx \vec{F}_i(x) - \sum_j \zeta_{ij} \vec{v}_j(t) + \vec{R}_i$$

- If all d.o.f. are eliminated and we work only with space- and time-dependent densities→ mesoscopic dynamics.
- Continuum fluid dynamics → Navier-Stokes
- Dissipative Particle Dynamics → Solve Navier-Stokes equations using an ensemble of special particles.
- Possible to study systems on the μm range during hundreds of μs.

QM/MM hybrid methods



QM subsystem embedded in MM system

A. Warshel & M. Levitt. *J. Mol. Biol.* 103: 227-249 (1976)

Hierarchical multiscale modelling

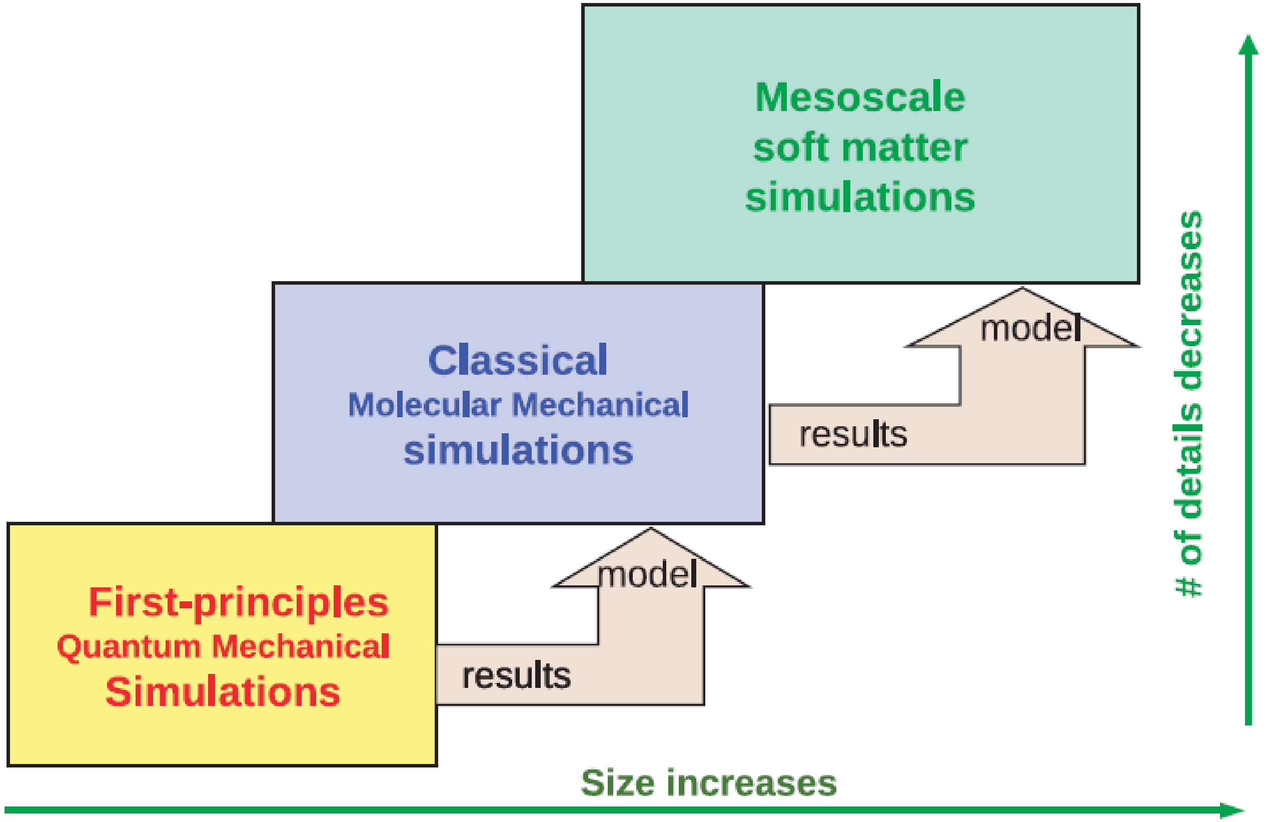
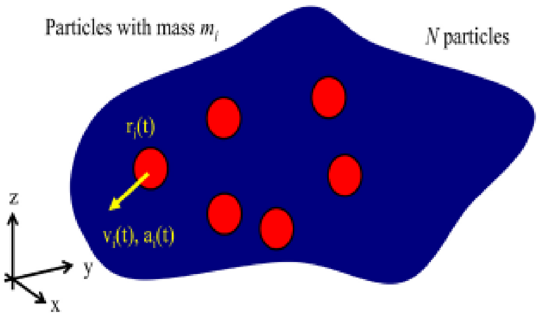


Fig. 1. Hierarchical multiscale approach: Results of simulations of a more detailed model are used to build a model for simulation on a larger scale. Within the process, the size of the system increases while the level of details decreases.

From Lyuvartsev et al., *J. Comput. Theor. Nanosci.* 6, 1-9 (2009)

Molecular Dynamics

Generate a dynamical trajectory by integrating Newton's equations of motion, with suitable initial and boundary conditions. We need a good way to determine the forces acting on each atom and a accurate numerical method to integrate the equations.



F = ma

$$m\frac{d^2 r_j}{dt^2} = -\nabla_{r_j} U(r_j) \qquad j = 1..N$$

Total energy of system

$$E = K + U$$

$$K = \frac{1}{2} m \sum_{j=1}^{N} v_j^2$$

$$U = U(r_j)$$

Coupled system N-body problem, no exact solution for N>2

System of coupled 2nd order nonlinear differential equations

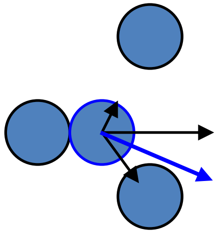
Solve by discretizing in time (spatial discretization given by "atom size")

Integrating the equation of motion

- Equations of motion in cartesian coordinates:

$$\frac{d\mathbf{r}_j}{dt} = \frac{\mathbf{p}_j}{m}$$

$$\frac{d\mathbf{p}_j}{dt} = \mathbf{F}_j$$



F: evaluate ½N(N–1) pairs

- **Important features of an integrator:**

  minimal need to compute forces (a very expensive calculation)
  good stability for large time steps
  good accuracy
  conserves energy and momentum
  time-reversible
  symplectic: conserves volume in phase space

Integrating the equation of motion

Discretize in time (n steps), with time step Δt:

$$r_i(t_0) \rightarrow r_i(t_0+\Delta t) \rightarrow r_i(t_0+2\Delta t) \rightarrow \ldots \rightarrow r_i(t_0+n\Delta t)$$

Simplest solution is to use a Taylor expansion:

$$r_i(t_0+\Delta t) = r_i(t_0) + v_i(t_0)\Delta t + \frac{1}{2}a_i(t_0)\Delta t^2 + O(\Delta t^3)$$

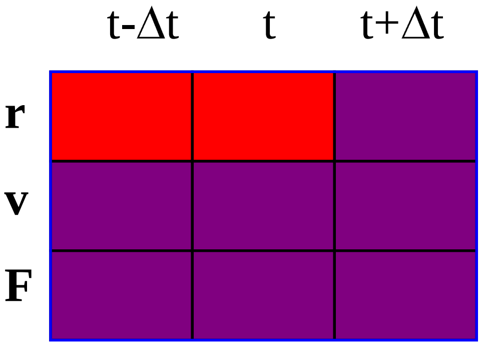
$$v_i(t_0+\Delta t) = v_i(t_0) + a_i(t_0)\Delta t + O(\Delta t^2)$$

But it does not work too well → **Unstable and inaccurate!**

Integrating the equation of motion: Verlet algorithm

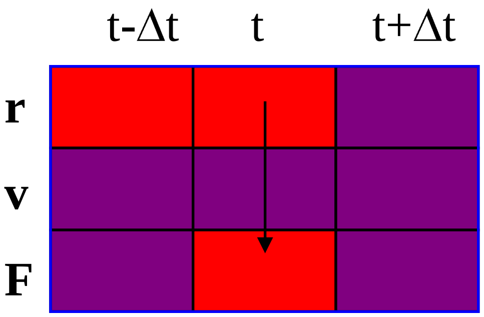
$$r_i(t_0+\Delta t)=r_i(t_0)+v_i(t_0)\Delta t+\frac{1}{2}a_i(t_0)\Delta t^2+\frac{1}{3!}\dddot{r}_i(t_0)\Delta t^3+O(\Delta t^4)$$

$$r_i(t_0-\Delta t)=r_i(t_0)-v_i(t_0)\Delta t+\frac{1}{2}a_i(t_0)\Delta t^2-\frac{1}{3!}\dddot{r}_i(t_0)\Delta t^3+O(\Delta t^4)$$



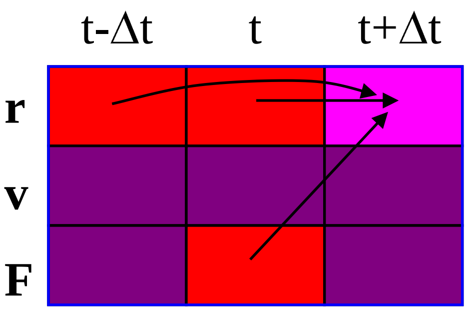
$$r_i(t_0+\Delta t)+r_i(t_0-\Delta t)=2r_i(t_0)+a_i(t_0)\Delta t^2+O(\Delta t^4)$$

Error is O(Δt^4) and it does not use the velocities.
They can be derived as:



$$r_i(t_0+\Delta t)-r_i(t_0-\Delta t)=2v_i(t_0)+O(\Delta t^3)$$

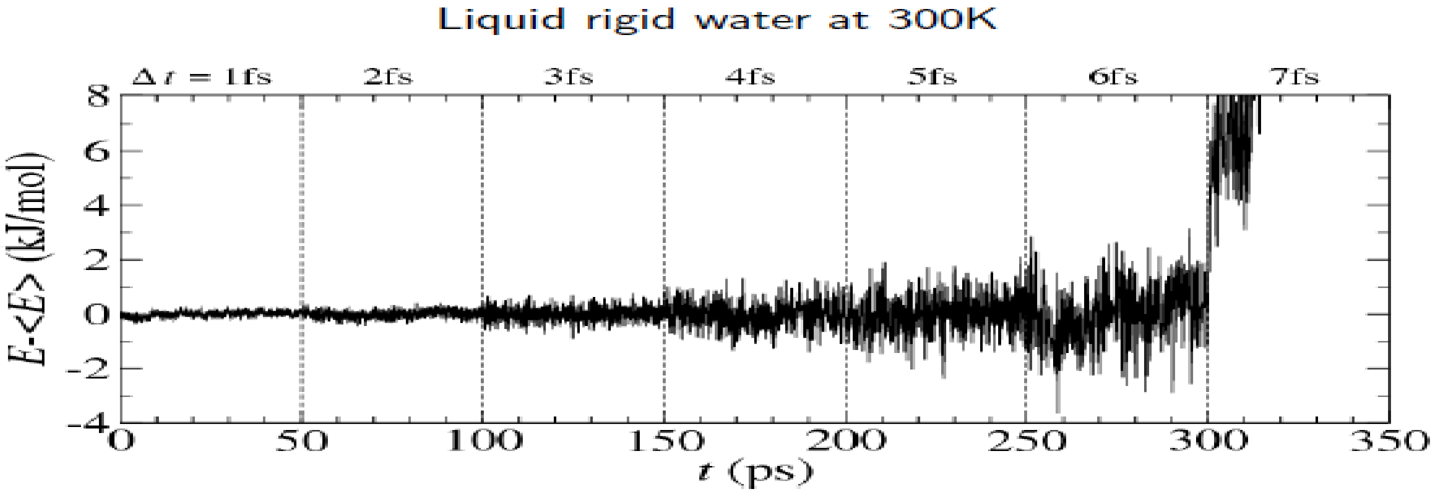
So $$v_i(t_0)=\frac{r_i(t_0+\Delta t)-r_i(t_0-\Delta t)}{2\Delta t}+O(\Delta t^2)$$



Integrating the equation of motion

- Choosing the time step:
  - too small: inefficient phase space sampling
  - too large: numerical instabilities → integrator divergence → MD crash

- Hints:
  - Δt should not be larger than the meantime between collisions
  - flexible molecules and rigid bonds, 2fs
  - flexible molecules and bonds, 1fs
  - total energy should be conserved → $\langle \delta E^2 \rangle^{1/2}/E < 10^{-4}$

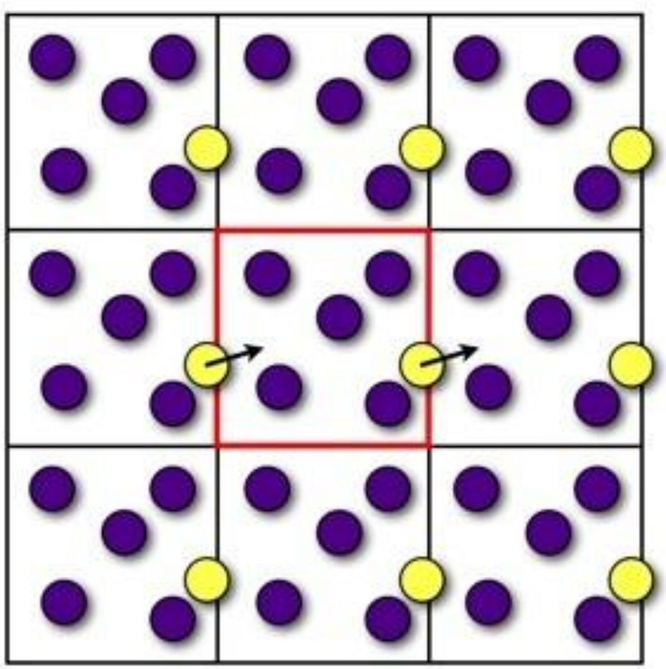


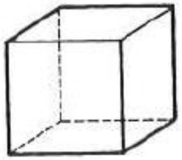
Periodic Boundary Conditions

- Enables bulk properties to be computed using a limited number of atoms

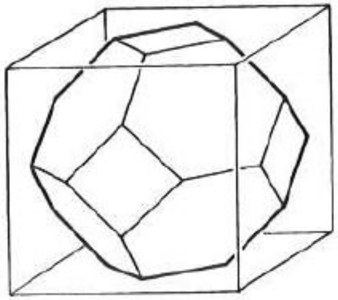
- Remove surface effects ~ $N^{-1/3}$ (49% for 1000 atoms)

- When an atom leaves the simulation cell, it is replaced by another with the same velocity, entering from the opposite cell face (number of atoms in the cell is conserved)

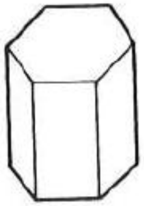




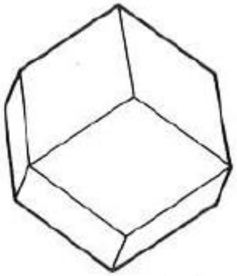
Cube



Truncated octahedron



Hexagonal prism

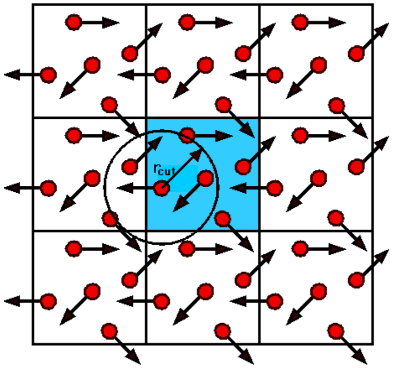


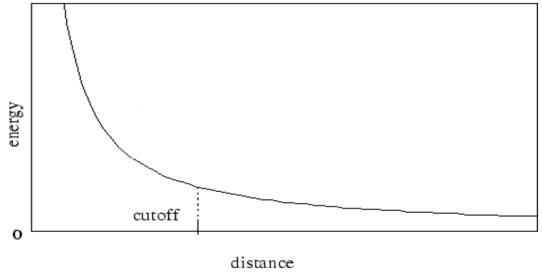
Rhombic dodecahedron

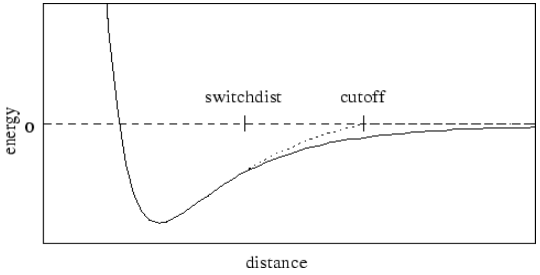
Non-bonded interactions

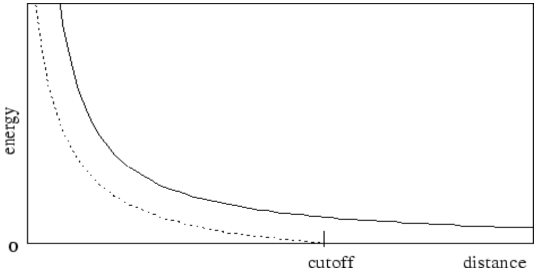
- Most time-consuming part of the simulation:
  - bonded interactions → O(N)
  - non-bonded interactions → O(N^2) → **time-consuming**

- Dealing with non-bonded interactions:
  - not feasible to include interactions with all images
  - minimum image convention:
    - an atom just sees the closest image of every other atom in the system,
  - use a non-bonded cutoff:
    - truncate the potential
    - consider interactions only inside a sphere of radius r_{cut}
    - discontinuity in force and energy calculation (truncated)



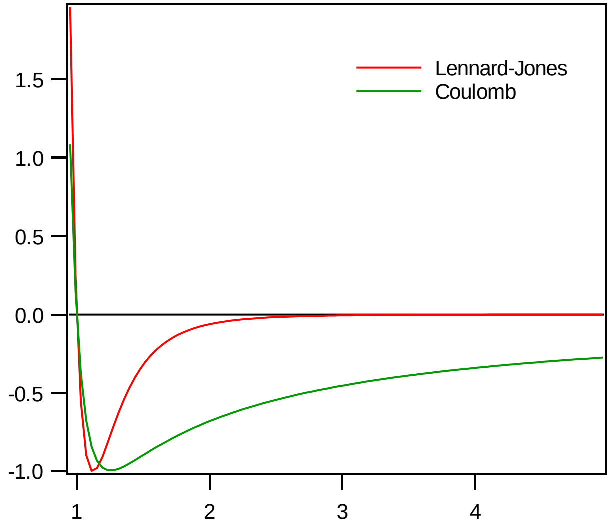






Electrostatic interactions

- Very long range interactions
  - far beyond primary cell

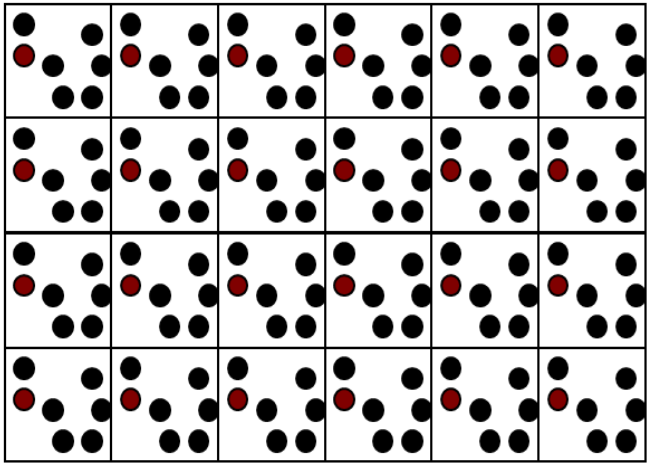


- 1/r does not die off as quickly as volume grows

$$\int_{r_c}^{\infty} \frac{1}{r} 4\pi r^2 dr = \infty$$

- Finite only because + and – contributions cancel

- Methods:
  - treat surroundings as dielectric continuum
  - full lattice sum → Ewald sum



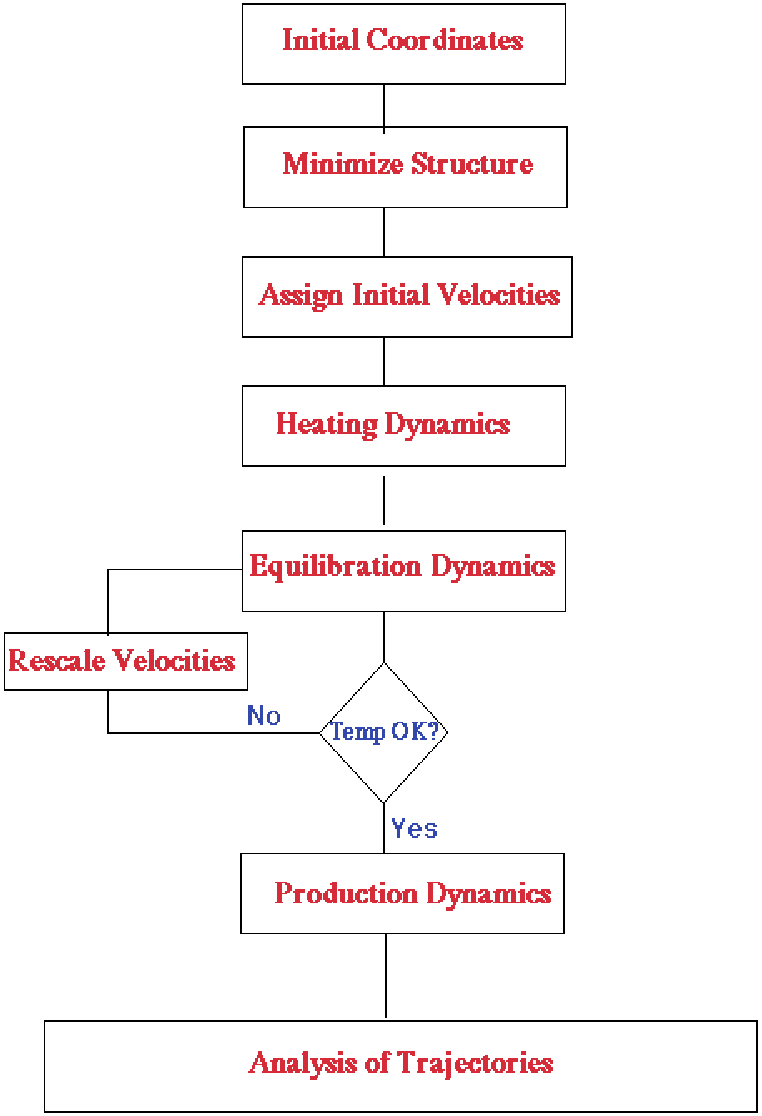
Thermodynamic ensembles

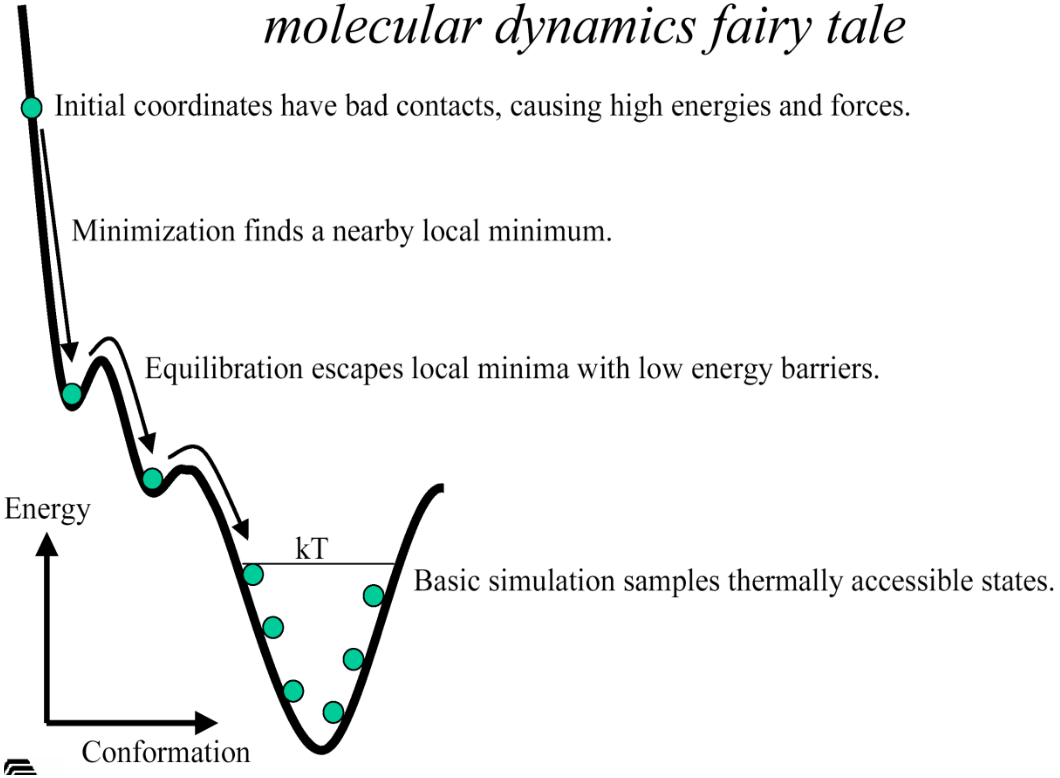
- The integration of the equations of motion keeps constant N, V, and E → **microcanonical** ensemble.
- But integration errors, force fluctuations and inconsistencies in the forces (e.g. generated by the cutoff) may cause slow drifts in the total energy.
- E=const, but not K and U, so systems not in equilibrium will go to equilibrium while the temperature changes.
- We may prefer to work at constant T or P or both to compare to experiment.

But we can modify the Lagrangian or couple the system to a heat or pressure bath:

- Several thermostats allow to do **NVT** simulations: velocity scaling, Berendsen, Andersen, Nosé-Hoover. But not all of them sample strictly the correct thermodynamic ensemble.
- There are also different barostats to do **NPT** simulations.

Protocol





molecular dynamics fairy tale

Initial coordinates have bad contacts, causing high energies and forces.

Minimization finds a nearby local minimum.

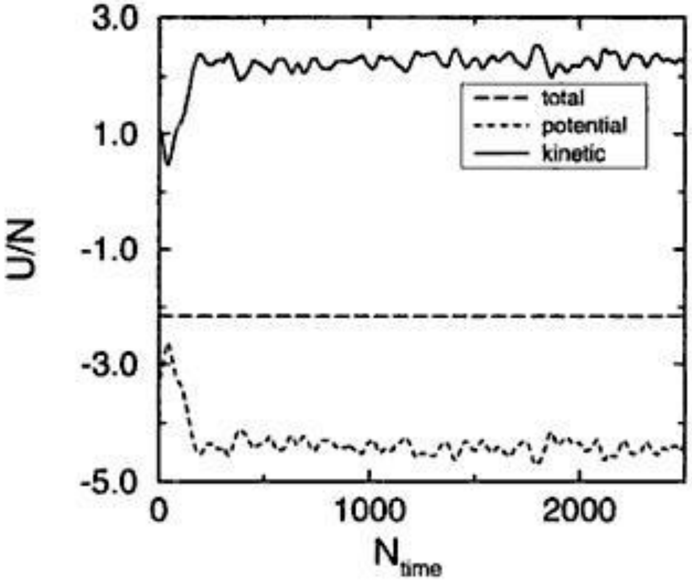
Equilibration escapes local minima with low energy barriers.

Basic simulation samples thermally accessible states.

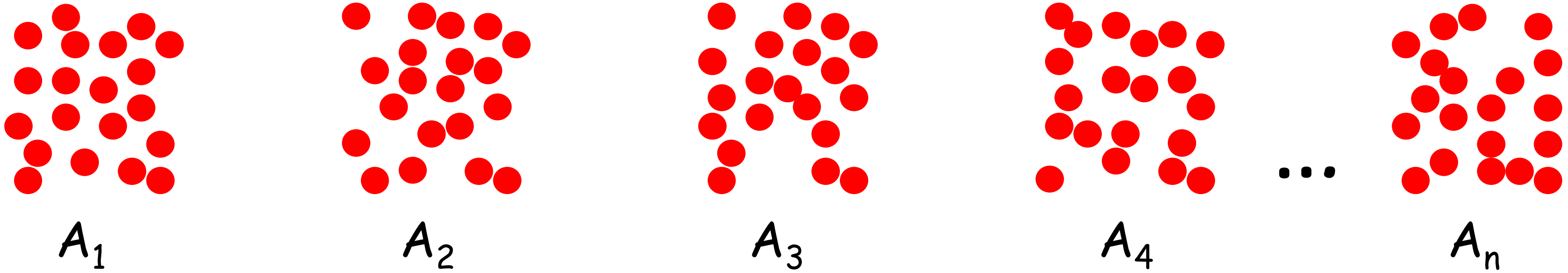
Protocol

To check:

- energy conservation
- no drift in temperature
- equilibration:
    - exchange between K and U
    - K, U, ρ converge
    - loss of long-range order for liquids
    - RMSD plateau for macromolecules
- production long enough for:
    - an efficient sampling
    - applying ergodic principle



Ergodic principle



- Ensemble = collection of large number of replicas:
  - each replica has the same macroscopic parameters (e.g. NVT)
  - differ microscopically → fluctuation
  - $\langle A \rangle$ = average over all replicas → **ensemble average**
- Molecular dynamics:
  - each time step generates a new configuration of a **single** replica
  - $\langle A \rangle$ = average over all configurations → **time average**
  - **Ergodic principle → ensemble average = time average**

$$\langle A \rangle = \int\int dp^N dr^N A(p^N, r^N) \rho(p^N, r^N) = \lim_{\tau \to +\infty} \frac{1}{\tau} \int_{t=0}^{\tau} A\left(p^N(t), r^N(t)\right) dt$$

Extracting the results: Thermodynamics properties

The main output of our simulation software will be a **trajectory** of our system: The positions (and velocities) of all the atoms as a function of simulation time. They can be used then to compute any property that can be expressed as a function of $r_i(t)$ or $v_i(t)$.

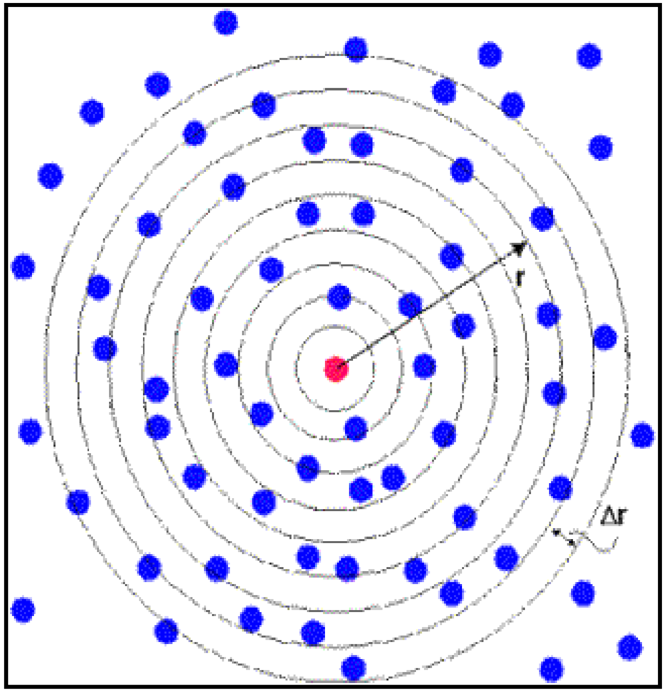
Running and average values for T, P, and the different contributions to the total energy are also given by most programs.

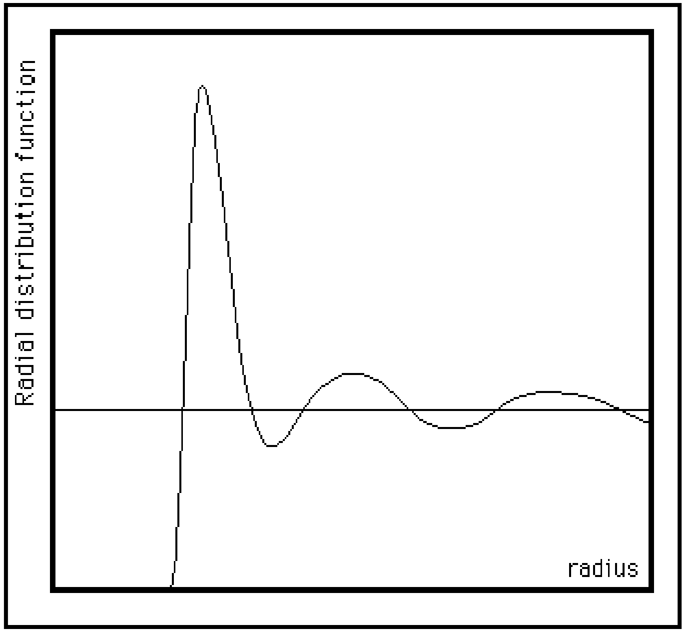
- Internal energy: $U = \langle E \rangle = \frac{1}{M} \sum_{i=1}^{M} E_i$
- Temperature: $\kappa = \sum_{i=1}^{N} \frac{\left| p_i \right|^2}{2m_i} = \frac{k_B T}{2} \left(3N - N_c \right)$
- Pressure: $P = \frac{1}{V} \left[N k_B T - \frac{1}{3} \sum_{i=1}^{N} \sum_{j=i+1}^{N} r_{ij} f_{ij} \right]$
- Heat capacity: $C_v = \left(\frac{\partial U}{\partial T} \right)_V = \frac{\langle E^2 \rangle - \langle E \rangle^2}{k_B T^2}$

- Useful to monitor the molecular dynamics and/or detect unconsistencies

Extracting the results: Structure

The radial distribution function (RDF) is a pair correlation function that describes how, on average, the atoms in a system are radially packed around each other.



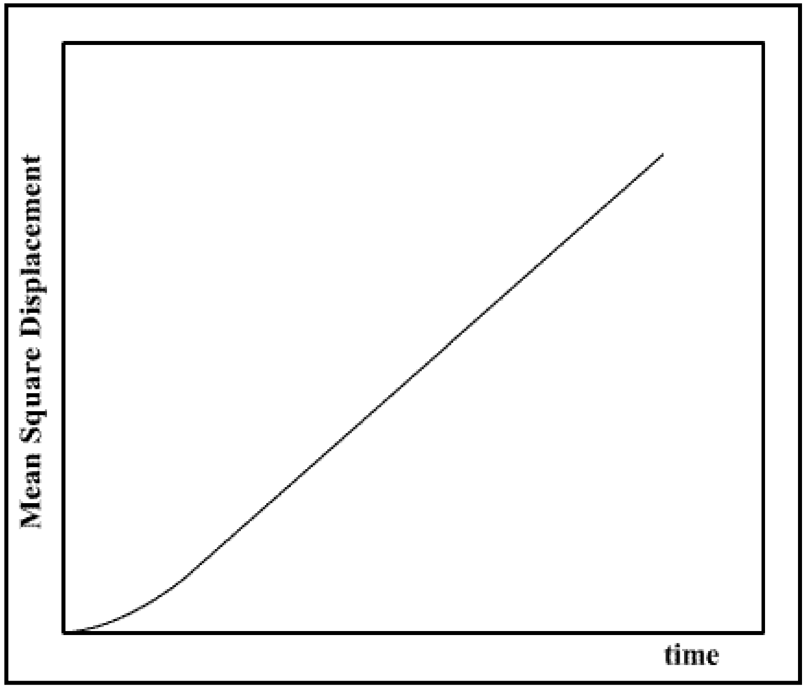


$$g(r)=n(r)/(\rho\ 4\pi\ \rho^2\Delta\rho)$$

Extracting the results: Self-diffusion

We can use Einstein expression, $\langle r^2 \rangle = 6 D t + C$, to determine the diffusion rate from the slope of the mean square displacement.

m.s.d. $\langle u^2(t)\rangle = (1/N_m)\, \Sigma_j\, [r_j(t_0+t) - r_j(t_0)\,]^2$



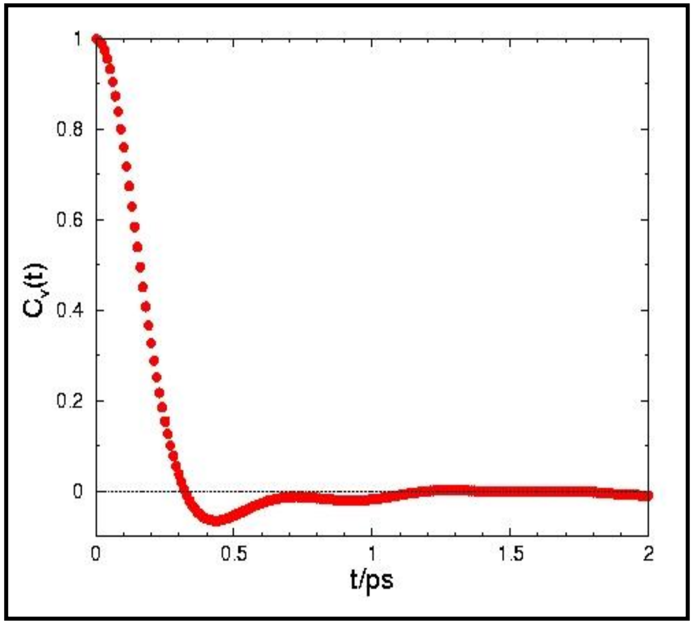
Note:

- Average over different time origins ⇒ errors increase with t.

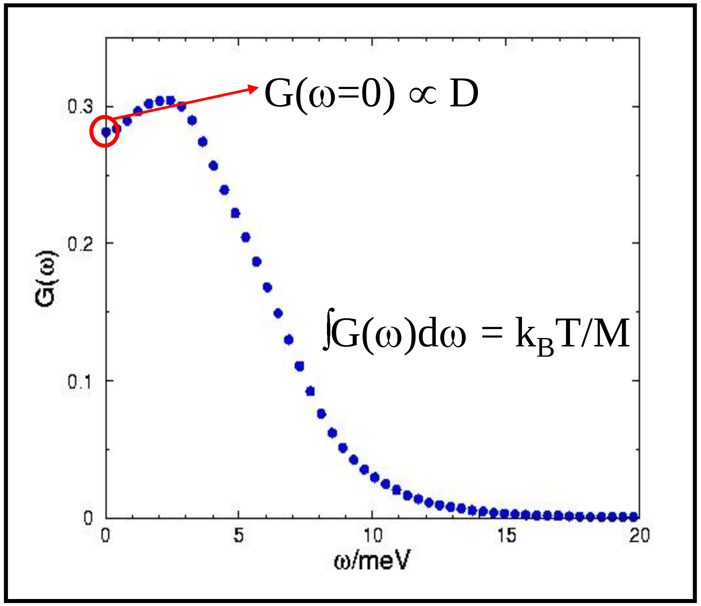
Extracting the results: Self-correlation functions

$$C(t) = (1/MN_m)\, \Sigma_j\, \Sigma_i\, A_i(t_j+t) \bullet A_i(t_j)$$

If $A_i = v_i \Rightarrow$ Velocity Autocorrelation Function (VACF)



FT



$G(\omega=0) \propto D$

$\int G(\omega)d\omega = k_BT/M$

Can be used to calculate D:

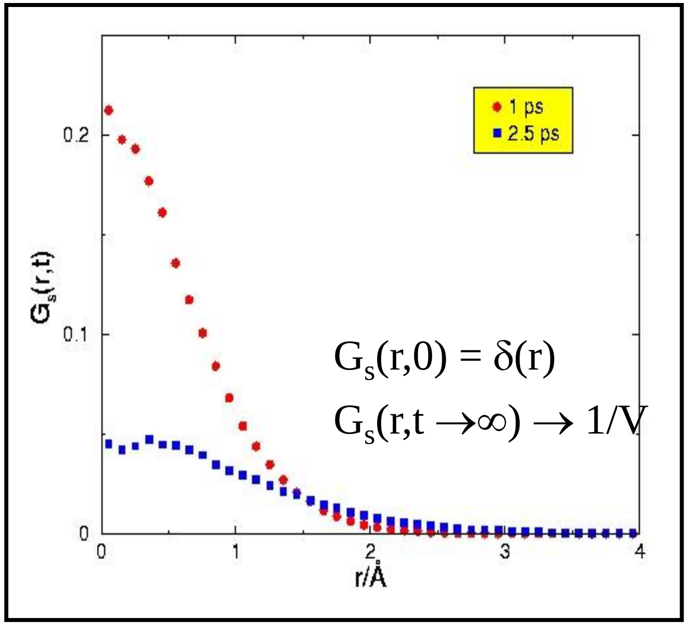
$$D = \int C_v(t)dt / 3$$

Density of States

Extracting the results: Van Hove correlation functions

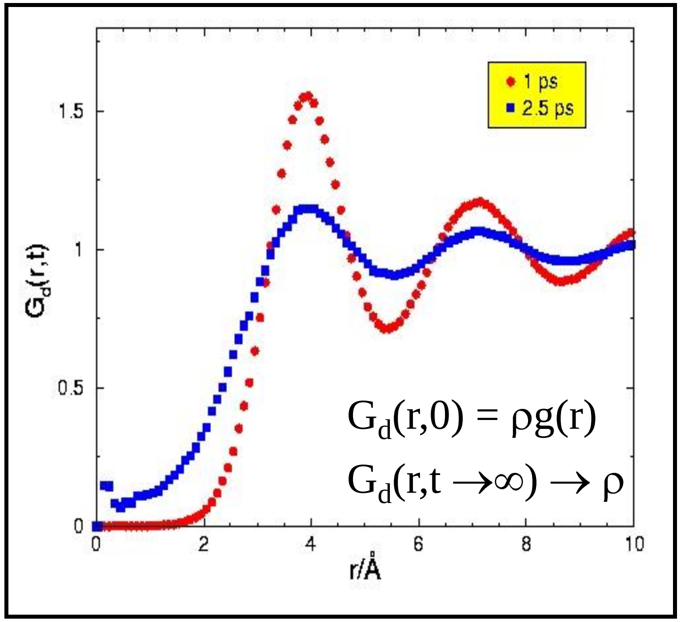
Van Hove correlation function:

$$G(r,t) = (1/N_m) \langle \Sigma_{ij} \delta[r + r_i(0) - r_j(t)] \rangle = G_s(r,t) + G_d(r,t)$$



$G_s(r,0) = \delta(r)$

$G_s(r,t \rightarrow \infty) \rightarrow 1/V$



$G_d(r,0) = \rho g(r)$

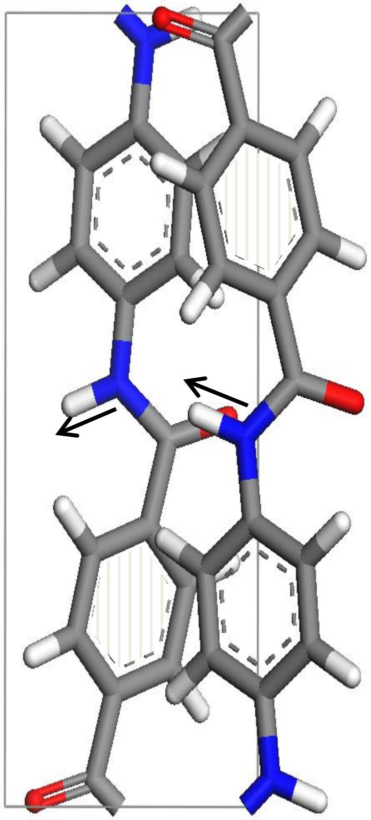
$G_d(r,t \rightarrow \infty) \rightarrow \rho$

FT in space of G(r,t) gives the intermediate scattering function, F(k,t), and its FT in space and time the dynamic structure factor, S(k,ω), which can be measured with scattering techniques (in particular, neutron scattering).

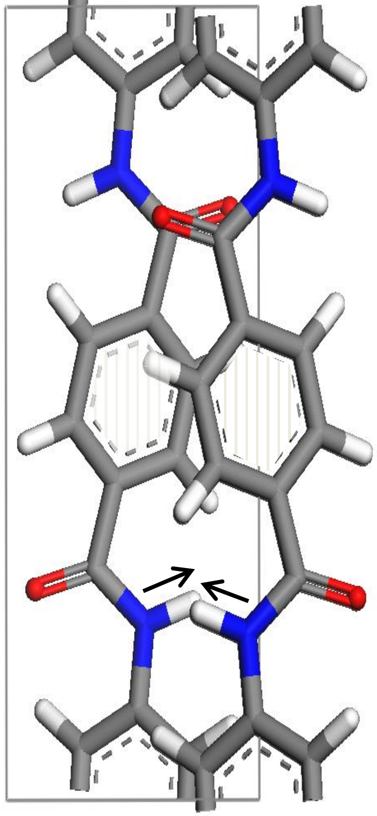
SOME EXAMPLES

Finding the right structure of Kevlar fibers

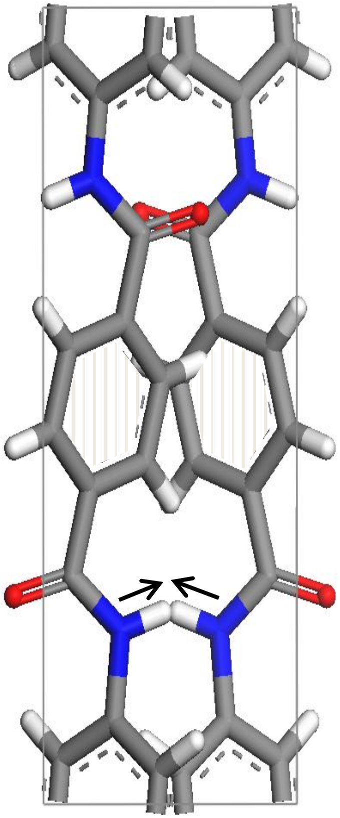
Northolt (*Pn*)



Northolt-translated (*Pn*)



Liu (*Pa*)

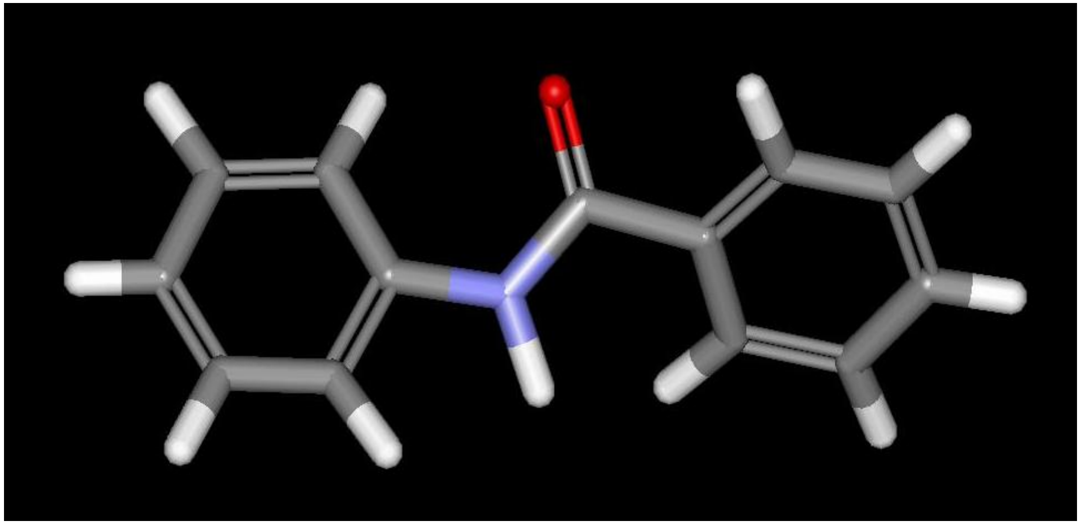


The (bio-) polymer; poly(p-phenylene terepthalamide)
Which is the correct structure?

Finding the right structure of Kevlar fibers

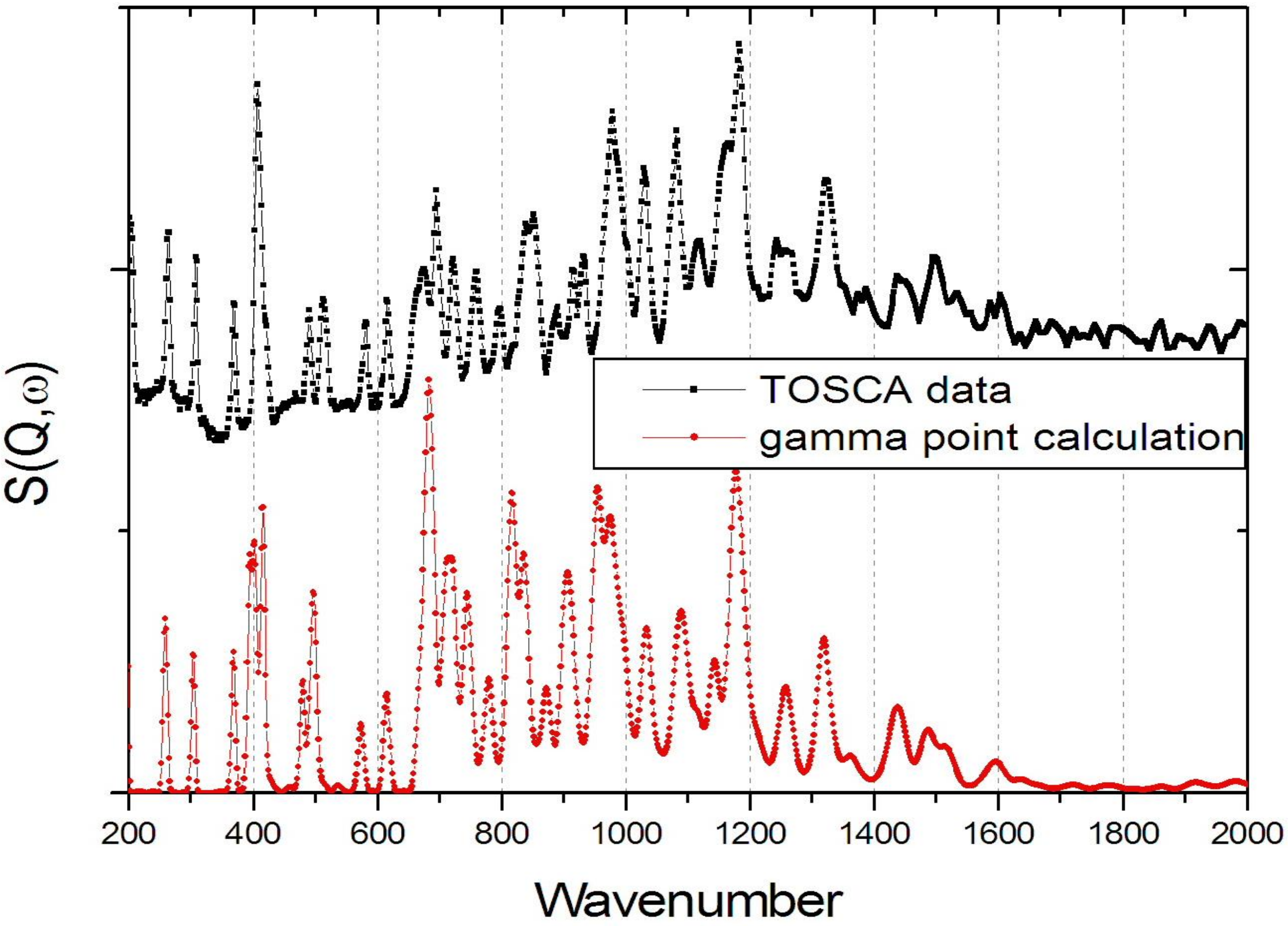
DFT simulations:

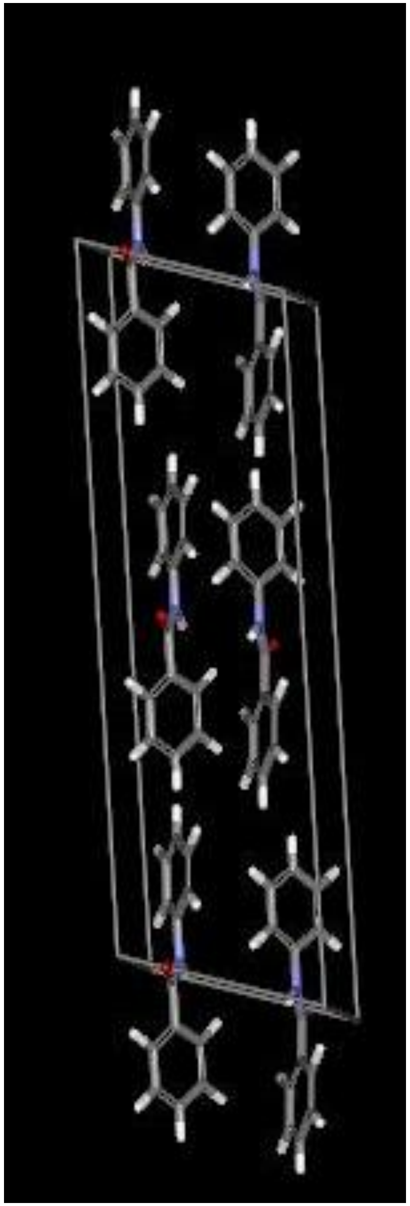
- Optimise crystal structure.
- Calculate force constants in supercell.
- Construct and diagonalise dynamical matrix for any k-vector in the Brillouin zone.
- Every k-vector has an associated set of eigenvalues and eigenvectors, i.e. mode assignments.



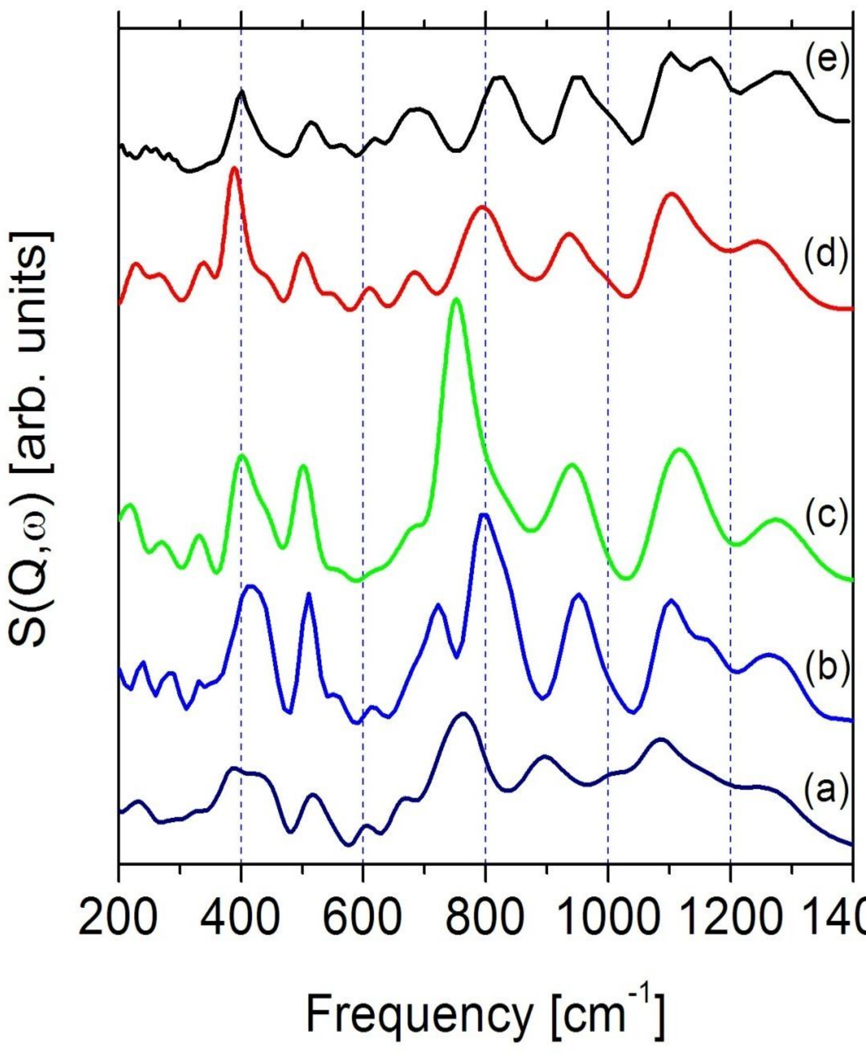
First validate the method on the monomer; benzanilide

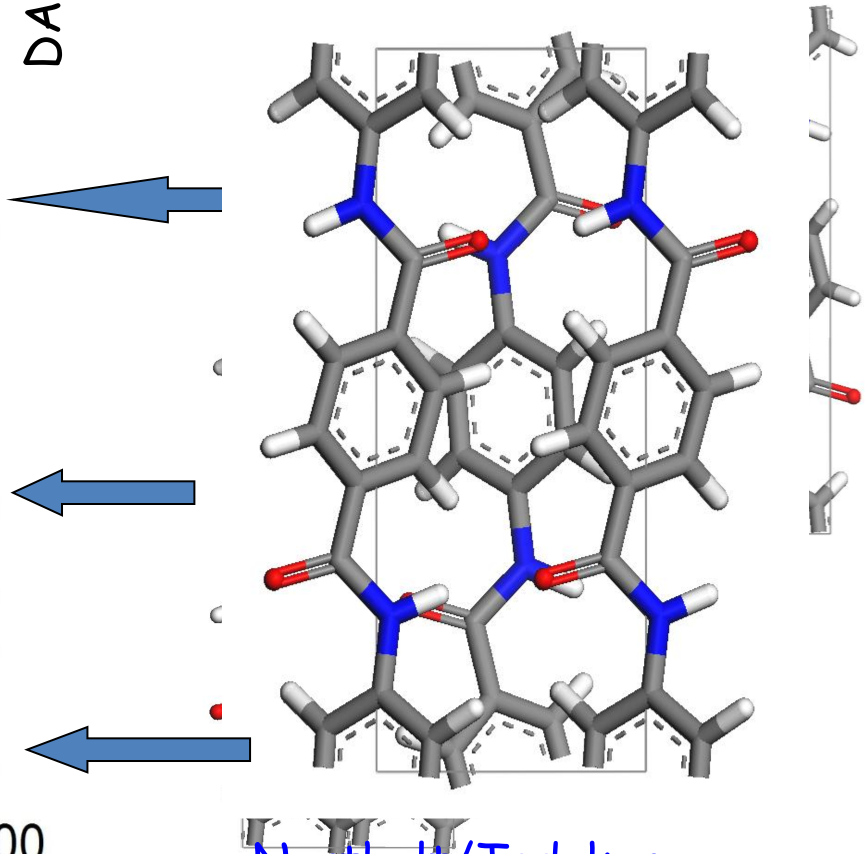
Finding the right structure of Kevlar fibers





Finding the right structure of Kevlar fibers





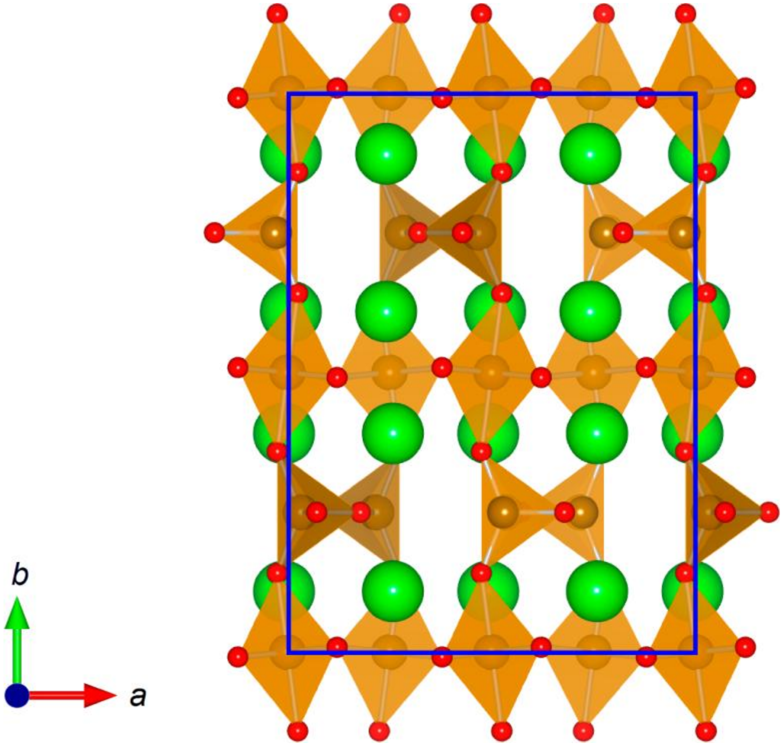
Northolt/Tadokoro

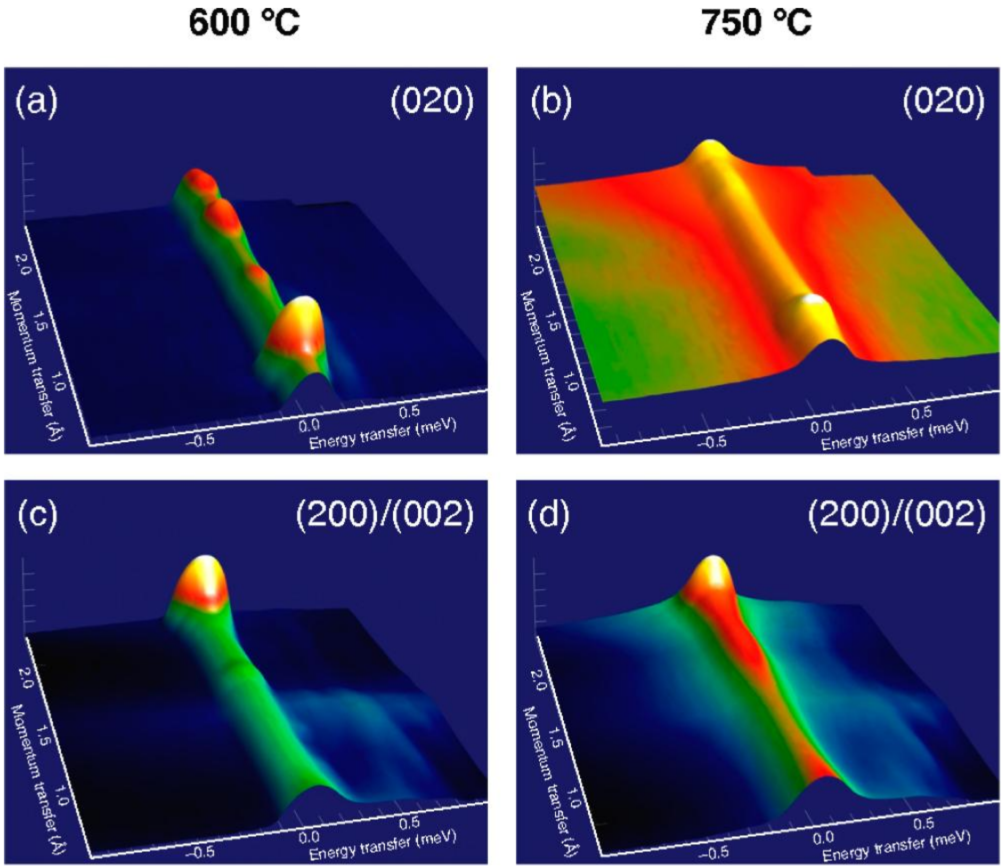
Pn or *P*2$_1$/*n*

Neutron Fibre diffraction *Pb*

Ion conduction dynamics in Sr2Fe2O5 Brownmillerite

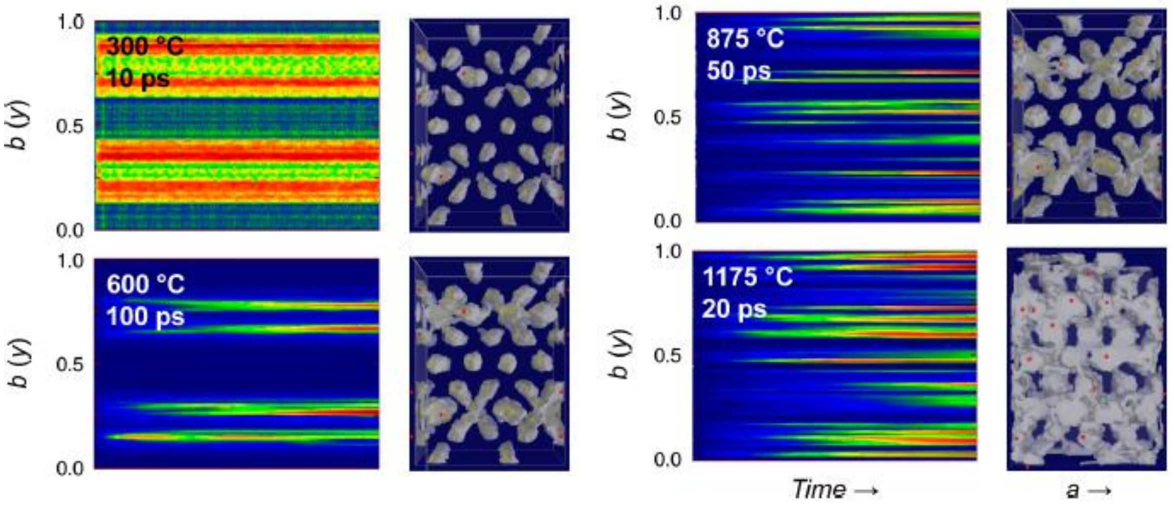
- Promising material for solid-state ion conducting applications
- Competitive oxide ion mobility at high T
- QENS data in a single crystal collected for two orientations
- QENS broadening at 750°C → HWHM ~ 0.16 meV, no marked Q dependence

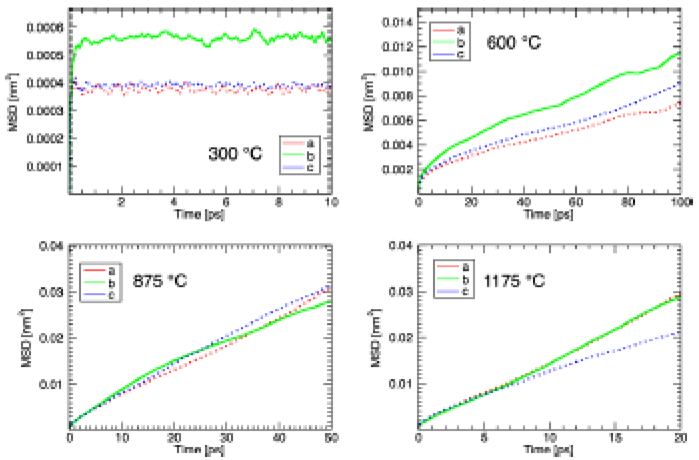




Ion conduction dynamics in Sr2Fe2O5 Brownmillerite

- AIMD simulations (VASP with GGA-PBE functional)
- 4 Icmm unit cells
- Longest simulation 150 ps @ 600°C, time-step 2 fs

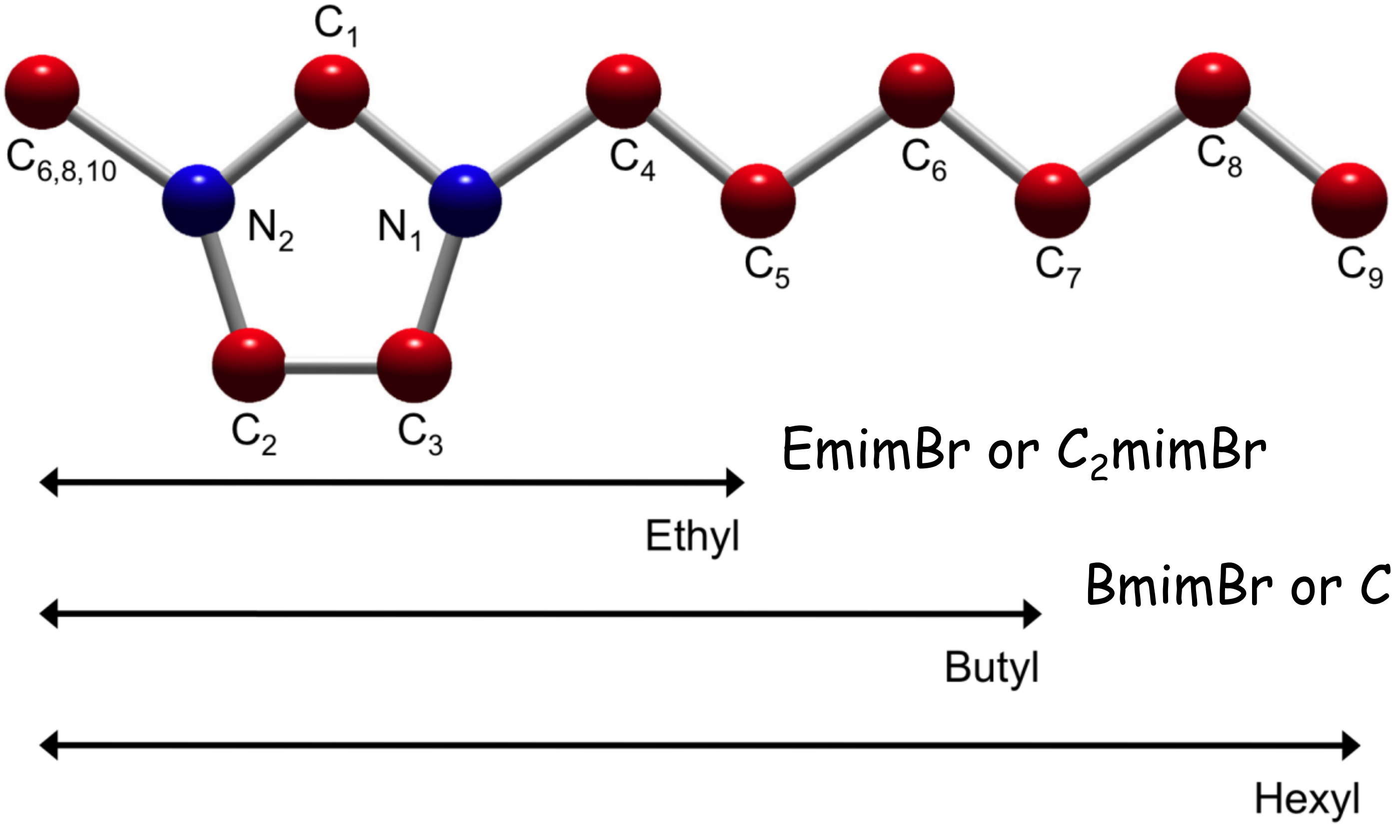




- Consistency between simulation and experiment.
- Largest motion in b direction, but at 875 °C becomes isotropic.
- Local brownmillerite-type vacancy order preserved beyond reported phase transition.
- Challenge: Longer simulations needed and enlarge simulation box.

Complex dynamics in room temperature ionic liquids

How does the microscopic structure and dynamics change with varying alkyl chain length?



EmimBr or C_2mimBr

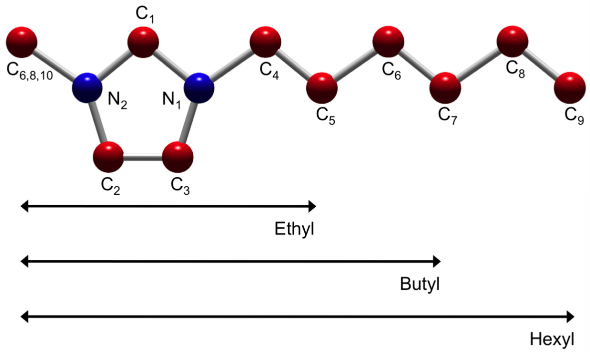
BmimBr or C_4mimBr

HmimBr or C_6mimBr

QENS analysis

$$F(Q,t) \approx DW \text{ x } T(Q,t) \times R(Q,t) \times L(Q,t)$$

$$S(Q,\omega) \approx \exp(-Q^2u^2)[T(Q,\omega) \otimes R(Q,\omega) \otimes L(Q,\omega)]$$



$$T(Q,\omega) \approx \mathcal{L}\,(\Gamma_T\,(Q) \approx DQ^2)$$

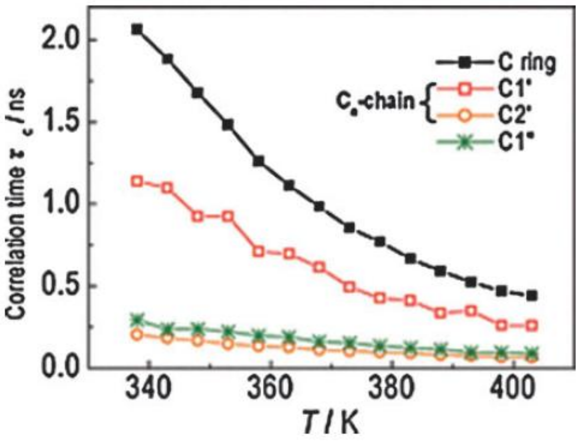
$$R(Q,\omega) \approx A_0^R + (1\text{-}A_0^R)\,\mathcal{L}\,(\Gamma_R)$$

$$L(Q,\omega) \approx A_0^L + (1\text{-}A_0^L)\,\mathcal{L}\,(\Gamma_L)$$

$$\begin{aligned} S(Q,\omega) \approx\; & A_0^R A_0^L\,\mathcal{L}\,(\Gamma_T) + \\ & (1-A_0^R)\,A_0^L\,\mathcal{L}\,(\Gamma_T+\Gamma_R) + \\ & A_0^R\,(1-A_0^L)\,\mathcal{L}\,(\Gamma_T+\Gamma_L) + \\ & (1-A_0^R)\,(1-A_0^L)\,\mathcal{L}\,(\Gamma_T+\Gamma_R+\Gamma_L) \end{aligned}$$

If $\Gamma_R \approx 0$ (MD, NMR Imanari 2010) then:

$$S(Q,\omega) \approx A_0^L\,\mathcal{L}\,(\Gamma_T) + (1-A_0^L)\,\mathcal{L}\,(\Gamma_T+\Gamma_L)$$



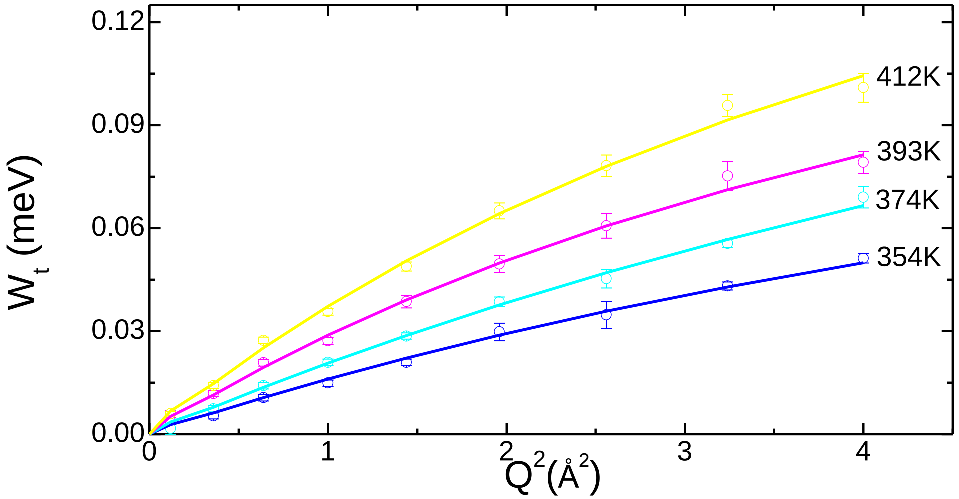
QENS: C_2mimBr dynamics

Data fitted with two lorentzians: 1 translational-like + 1 local-like

D follows Arrhenius law:
$D = D_0 \exp(-E_a / RT)$ with

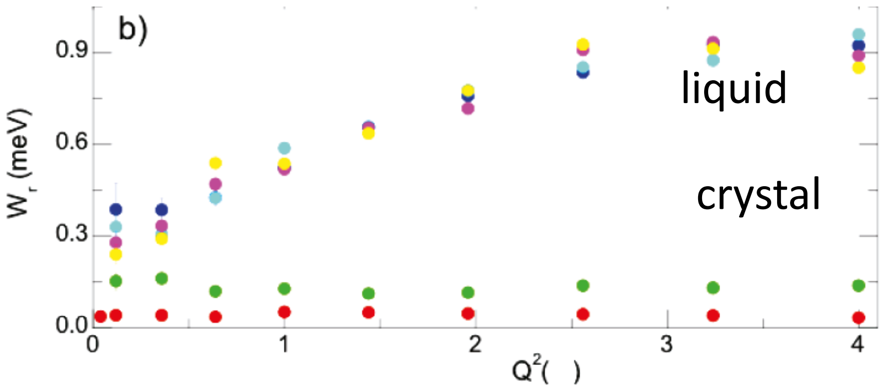
$D_0 = (1.7 \pm 0.8).10^{-7} m^2 \cdot s^{-1}$
$E_a = 19 \pm 2$ kJ·mol^{-1}

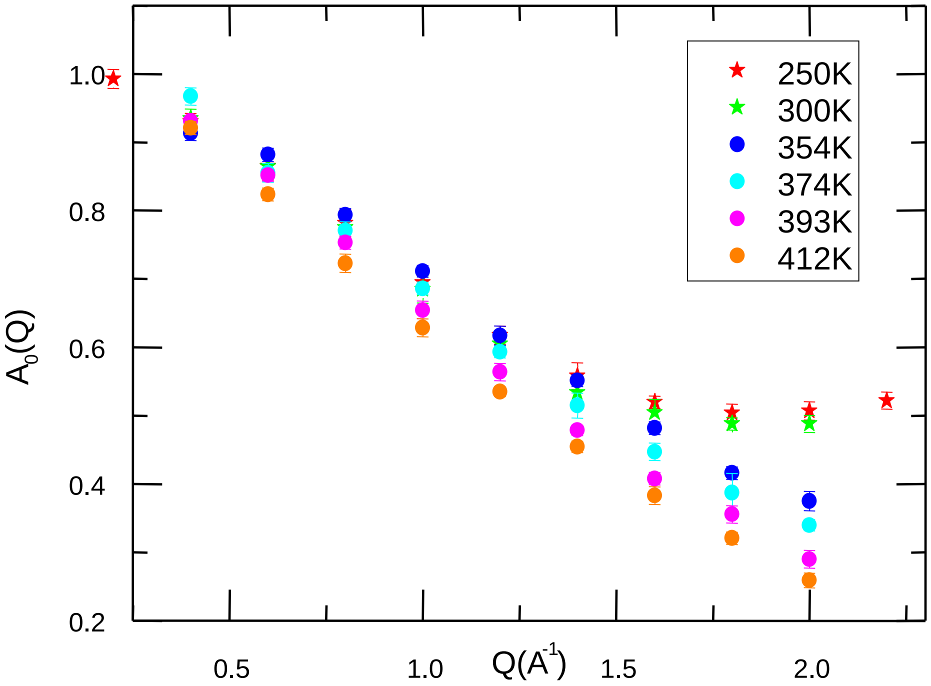
Reasonable agreement with NMR (Every, PCCP 2004), although D values 3-4 times larger.



T (K)	*D* (10^{-10} m^2·s^{-1})	τ_0 (ps)
353	2.7 ± 0.2	3.9 ± 0.6
373	3.4 ± 0.5	2.6 ± 0.4
392	5.1 ± 0.7	3.2 ± 0.2
412	6.6 ± 0.9	2.5 ± 0.2

C_2mimBr : Local dynamics

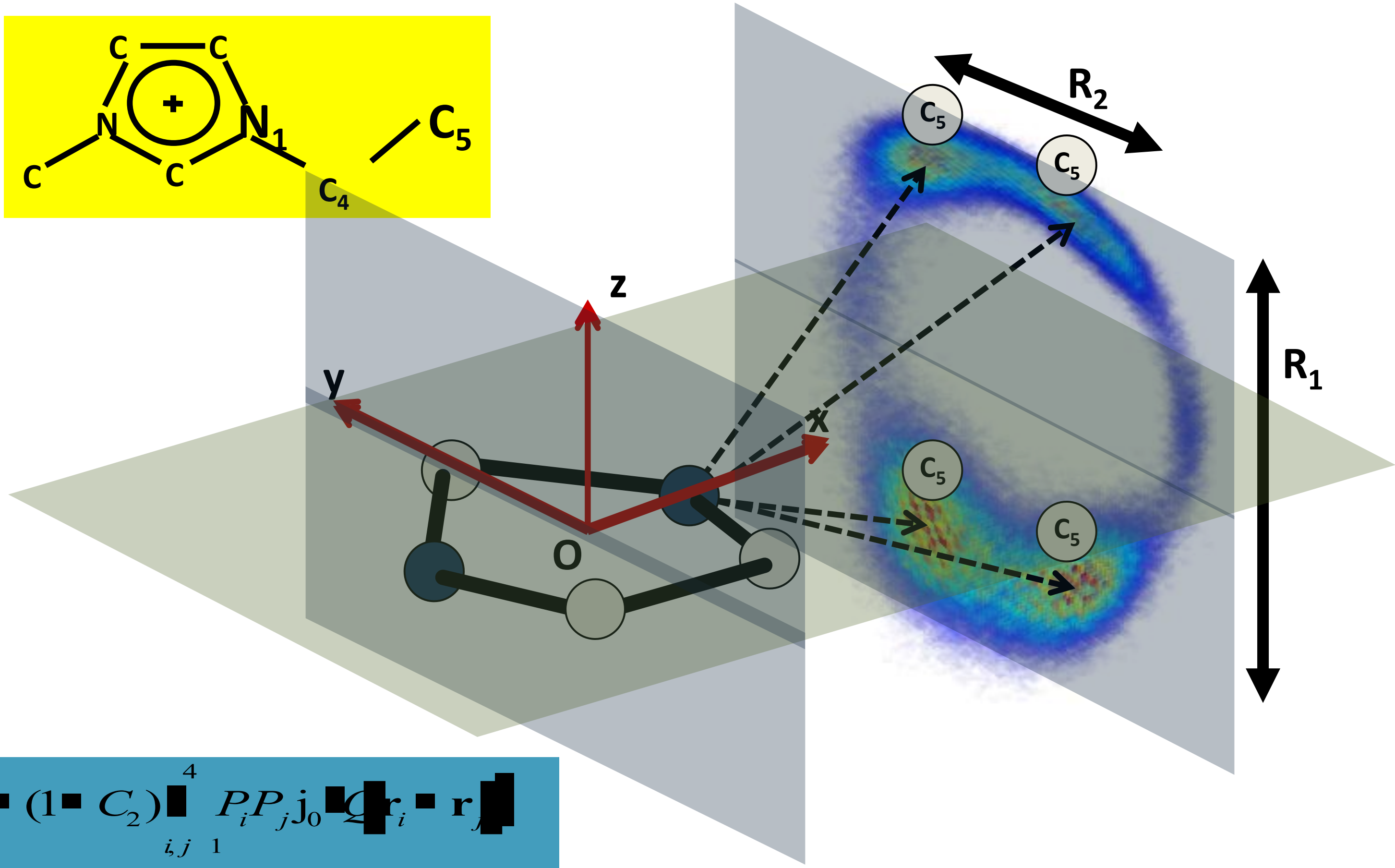




$$S(Q,\omega) = \exp\left(-Q^2\langle u^2\rangle\right)\left[A_0(Q)\delta(\omega) + \left(1 - A_0(Q)\right)L\left(\Gamma_l,\omega\right)\right]$$ (crystal)

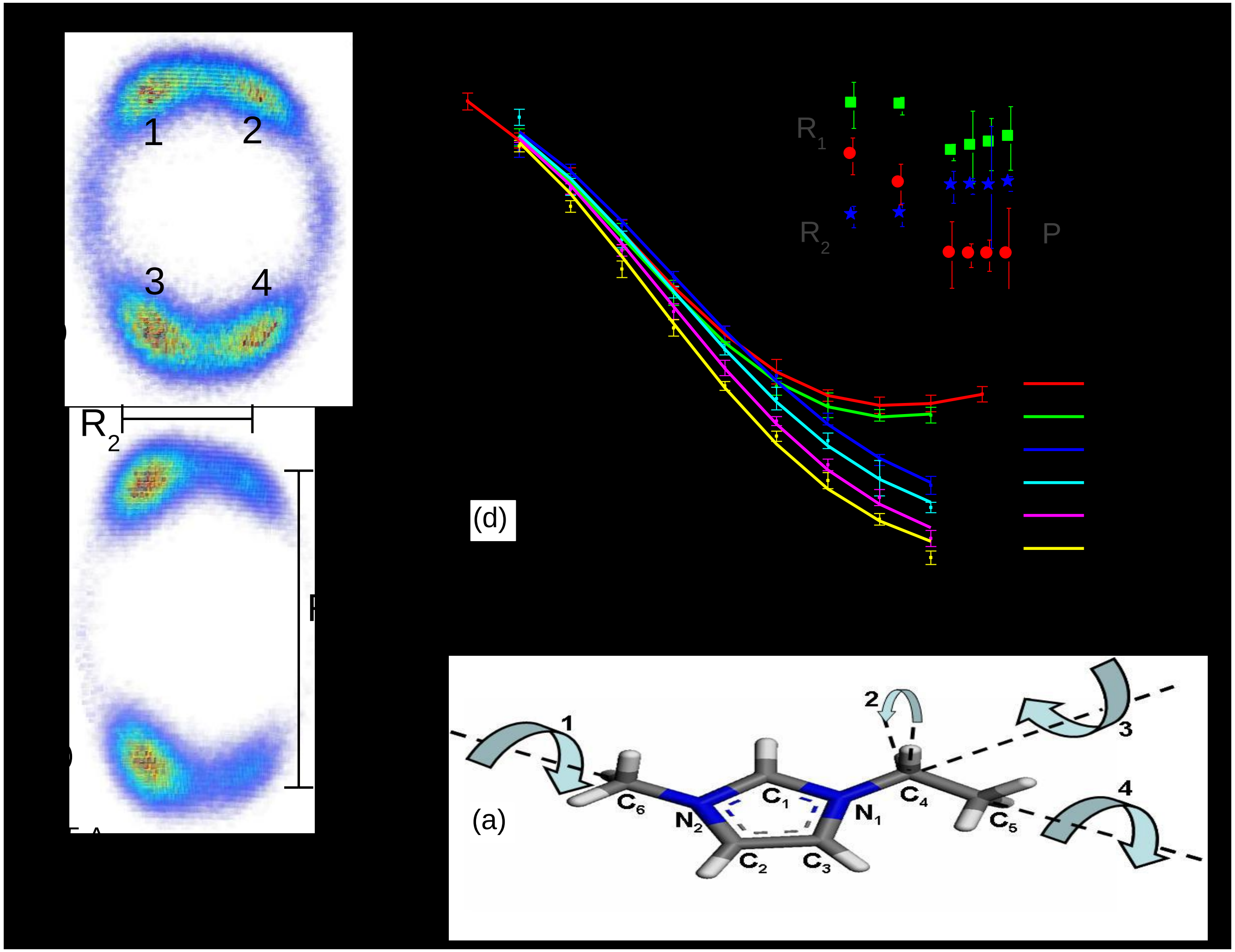
$$S(Q,\omega) = \exp\left(-Q^2\langle u^2\rangle\right)\left[A_0(Q)L\left(\Gamma_t,\omega\right) + \left(1 - A_0(Q)\right)L\left(\Gamma_{t+l},\omega\right)\right]$$ (liquid)

Motion of the ethyl chain from MD



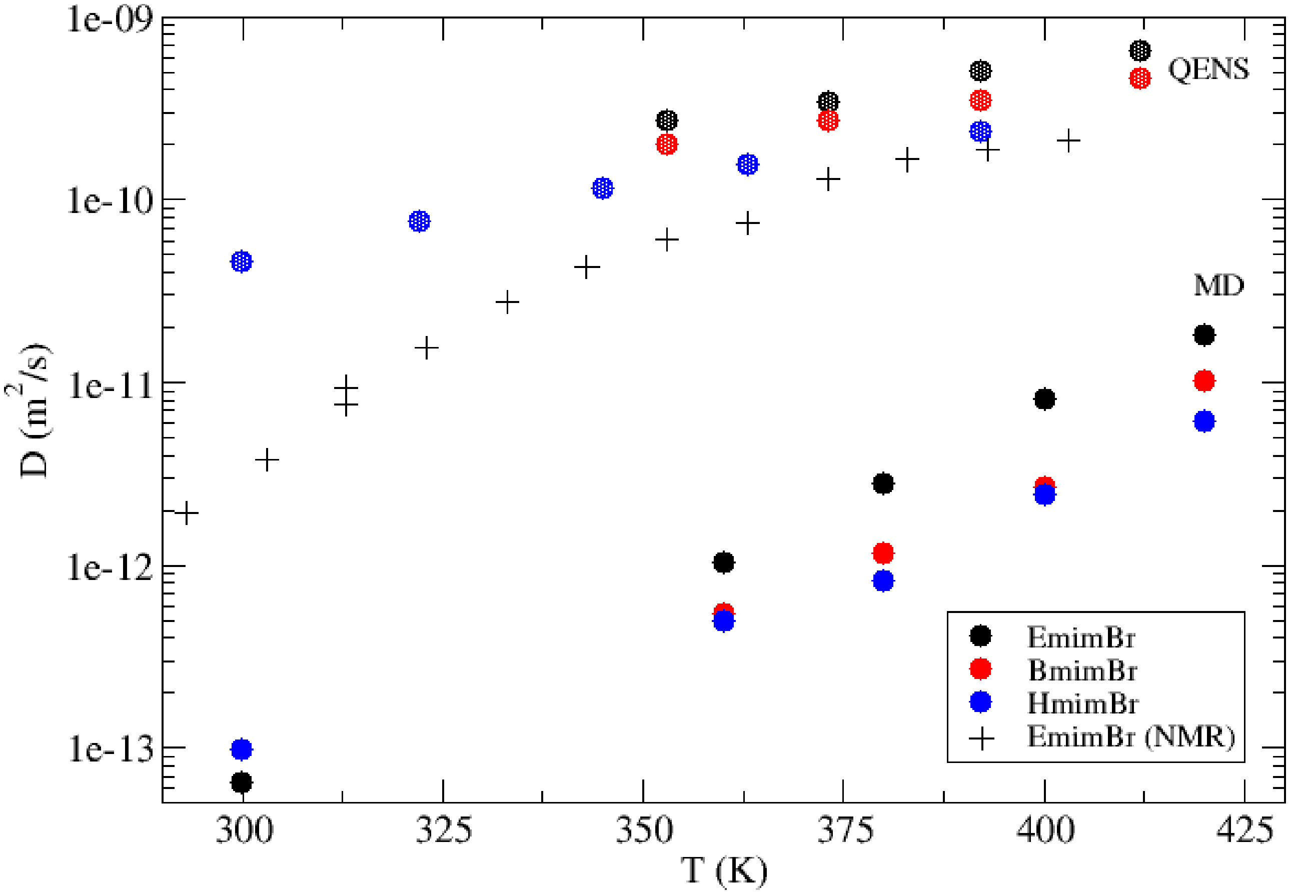
$$A_0(Q) = C_1 + (1 - C_2)\sum_{i,j=1}^{4} P_i P_j j_0(Q|\mathbf{r}_i - \mathbf{r}_j|)$$

QENS analysis using MD input



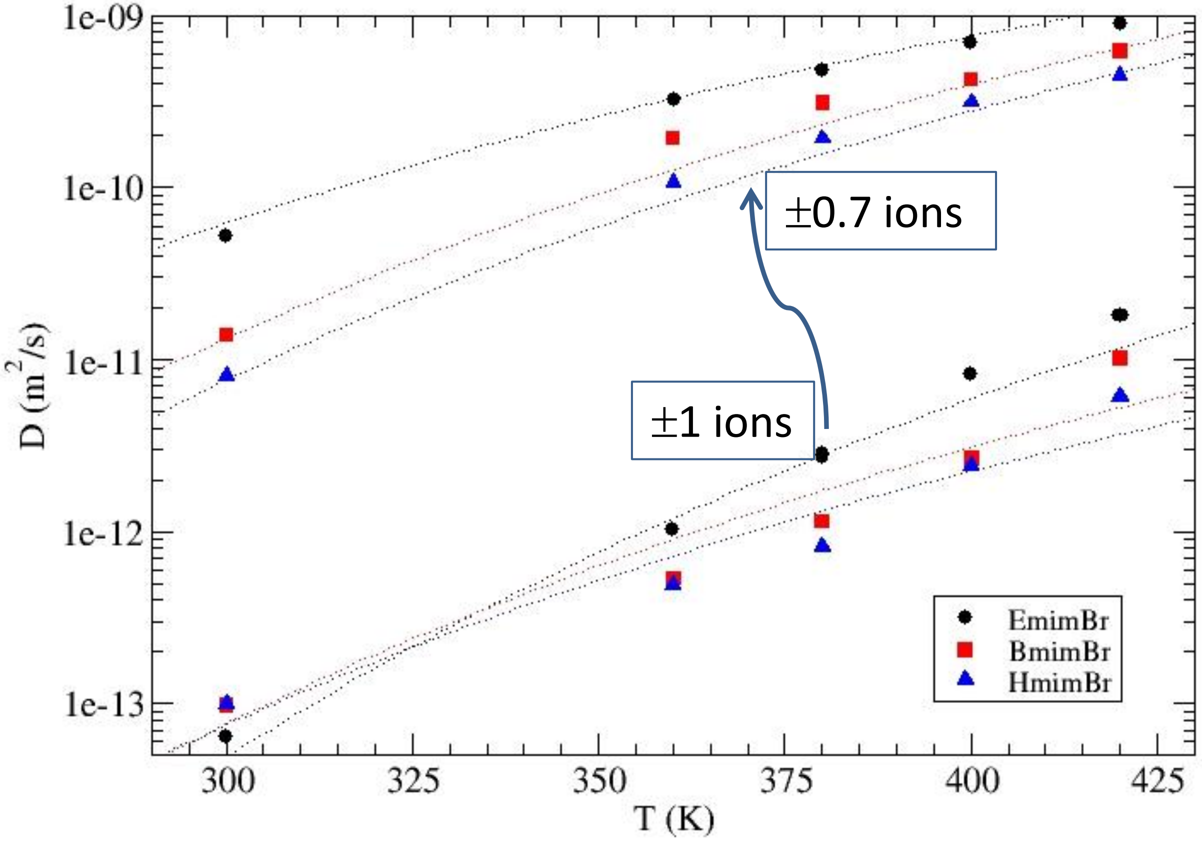
Aoun et al. , *J. Phys. Chem. Letters* **1**, 2503 (2010)

Self-diffusion coefficients



NMR data: Every et al., Phys. Chem. Chem. Phys. **6**, 1758 (2004)

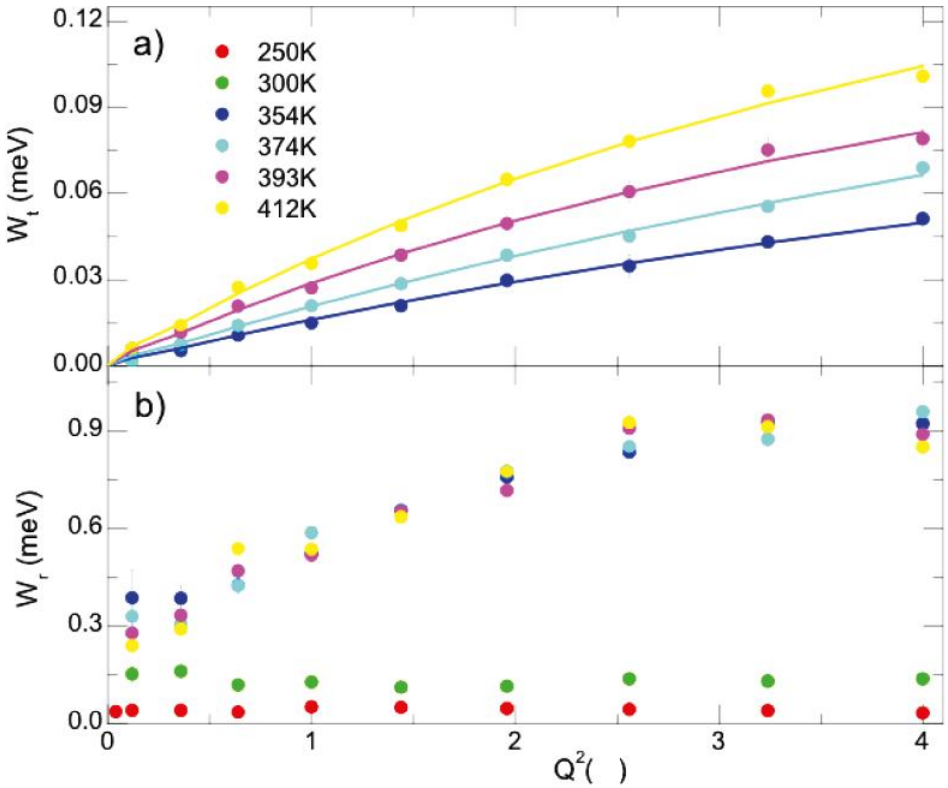
MD: Scaled *vs* full charges

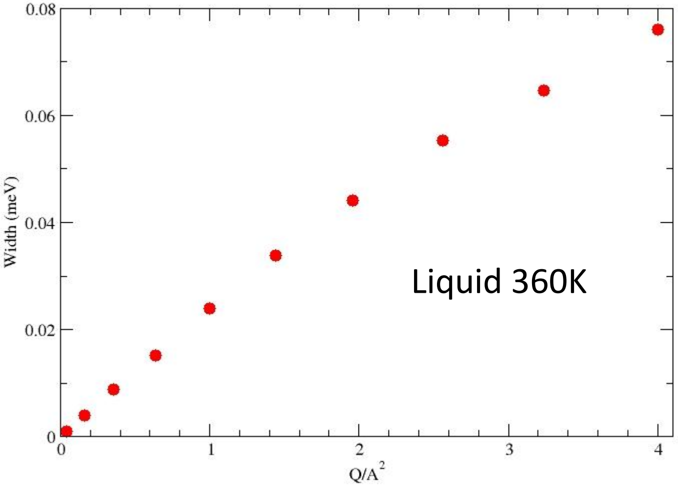


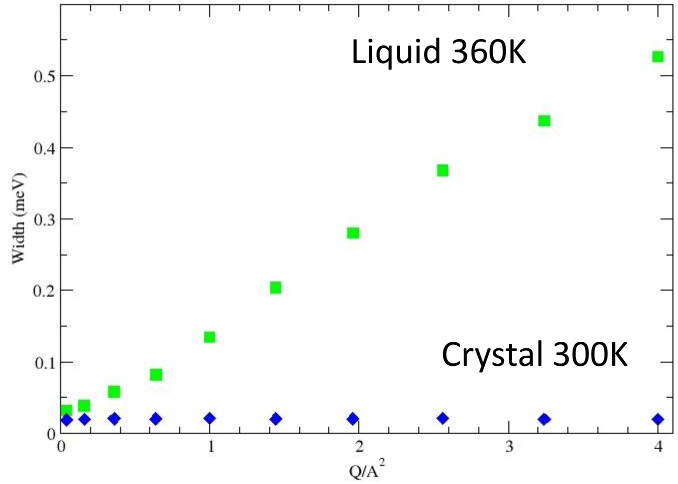
±0.7 ions

±1 ions

Quasielastic widths: Simulation vs Expt.

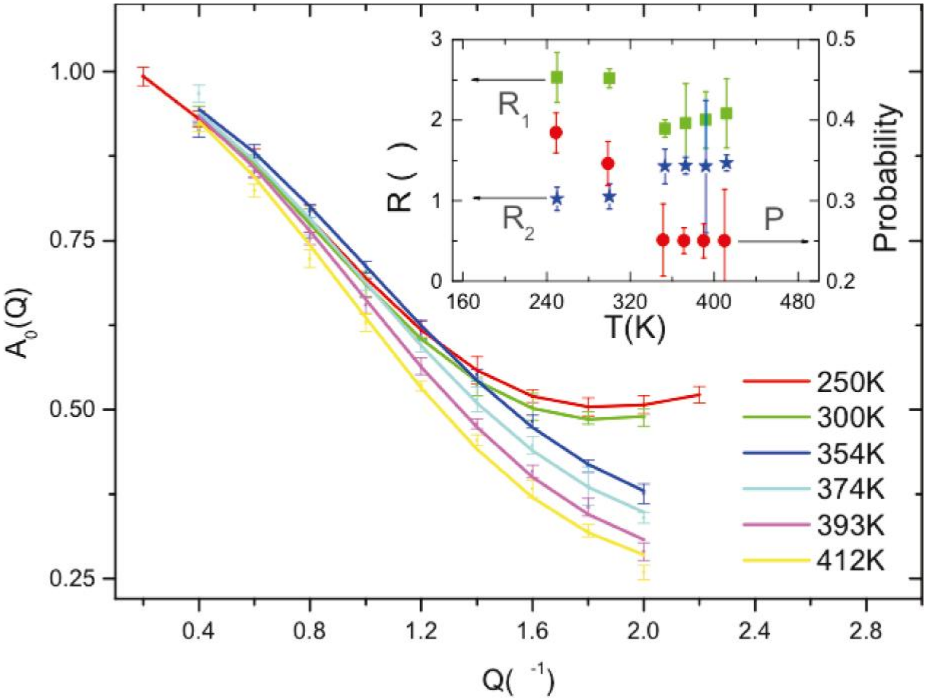


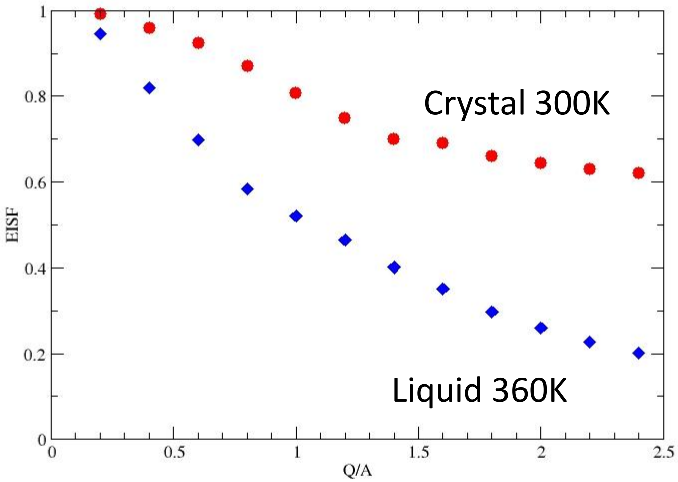




Simulation: D (from width of narrow line) = 4.9 x 10^{-10} m^2/s *vs* 3.2 x 10^{-10} m^2/s obtained directly from m.s.d.!

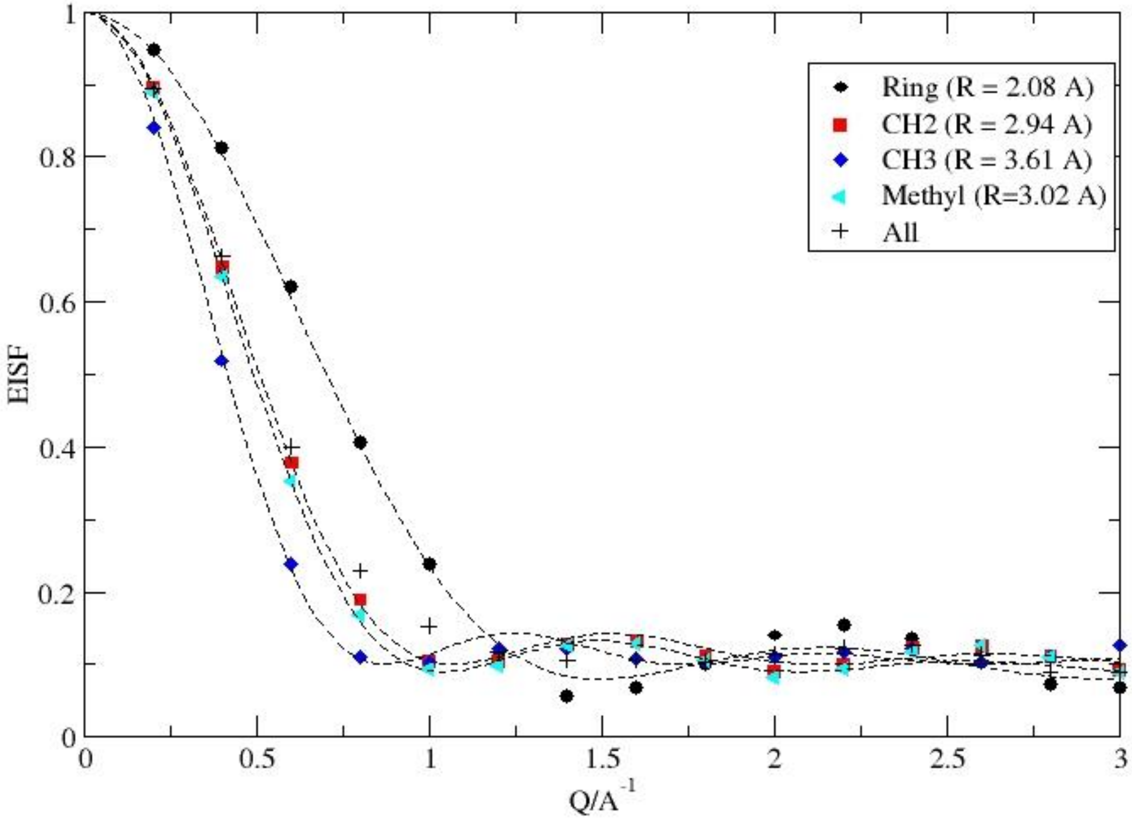
EISF: Simulation *vs* experiment





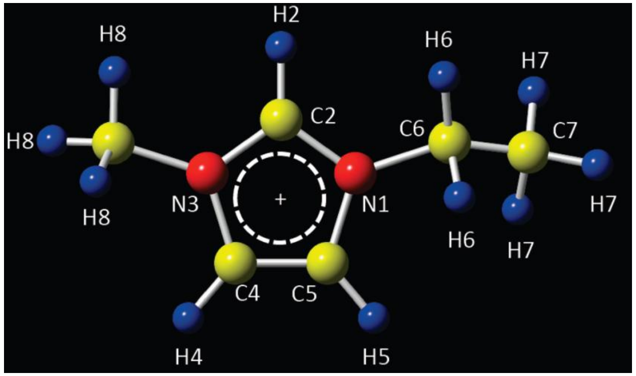
Qualitative or even semiquantitative agreement between experimental (fitted S(Q,ω)) and simulated (fitted F(Q,t) with equivalent model) widths and EISF's.

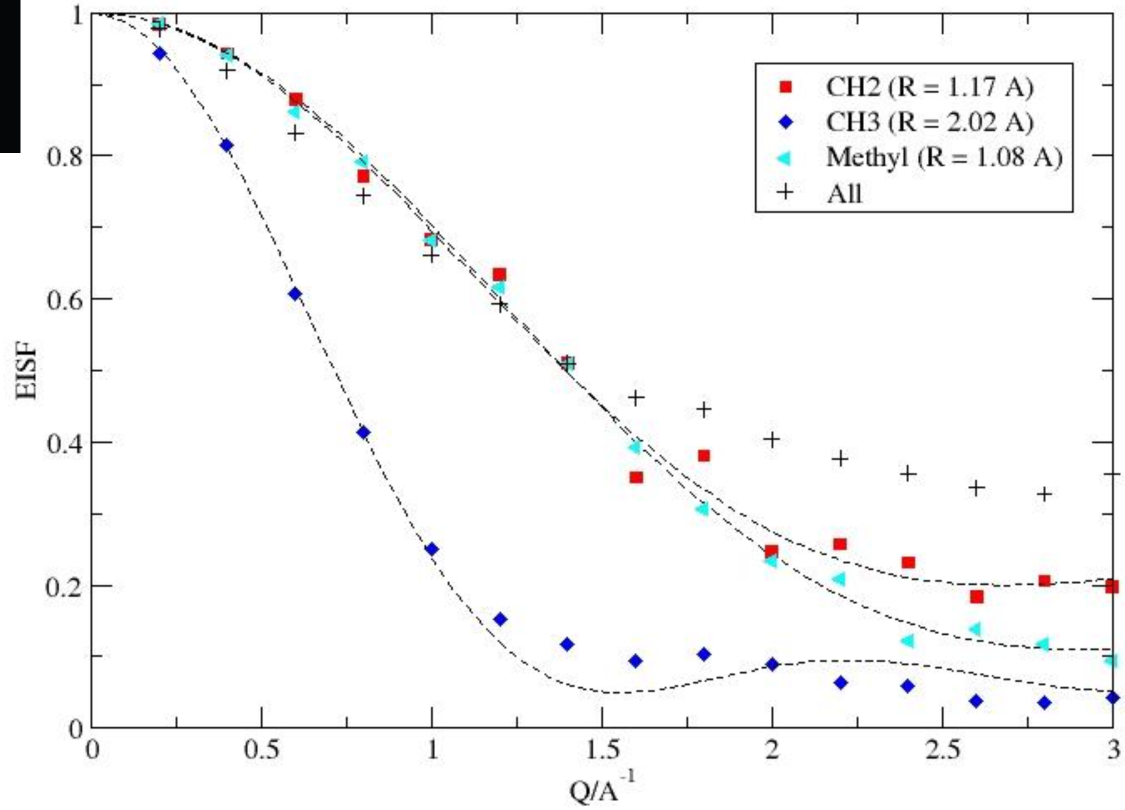
Global rotation F(Q,t) → EISF by groups



When looking to individual groups, reasonable agreement with model of diffusion on the surface of a sphere.

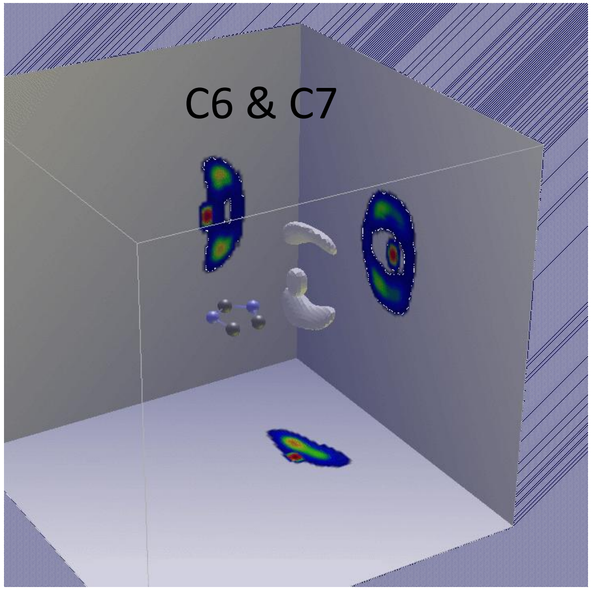
Simulated spectra: EISF for chain motions

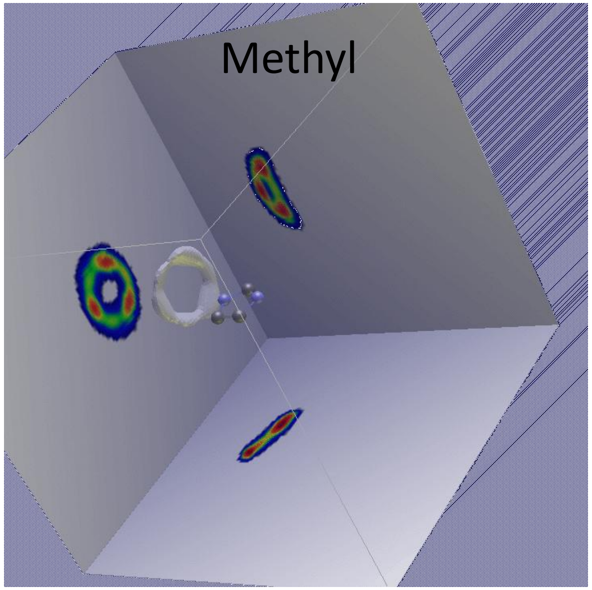


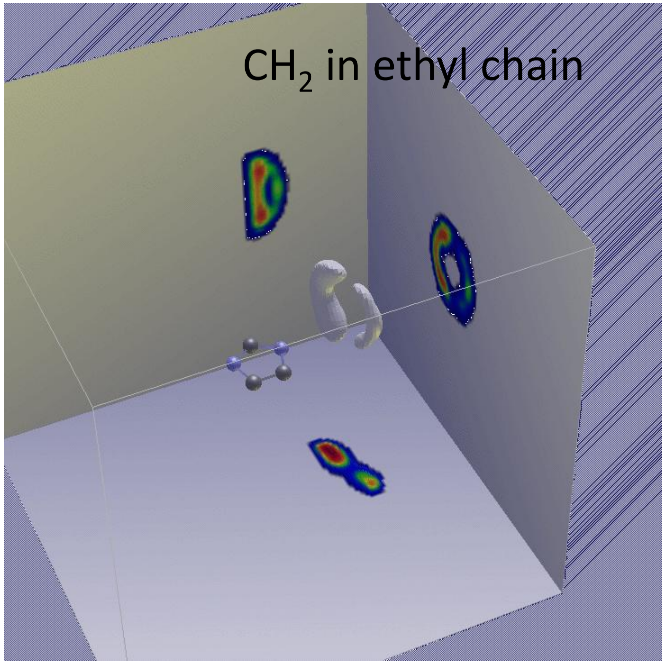


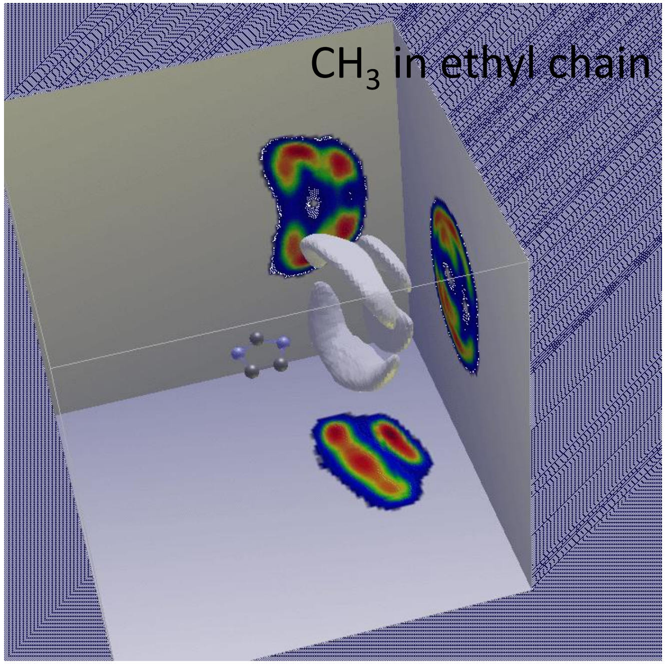
Possible to fit to model of rotation in a circle. But meaningful?

Local motions: Spatial distribution (crystal)


C6 & C7

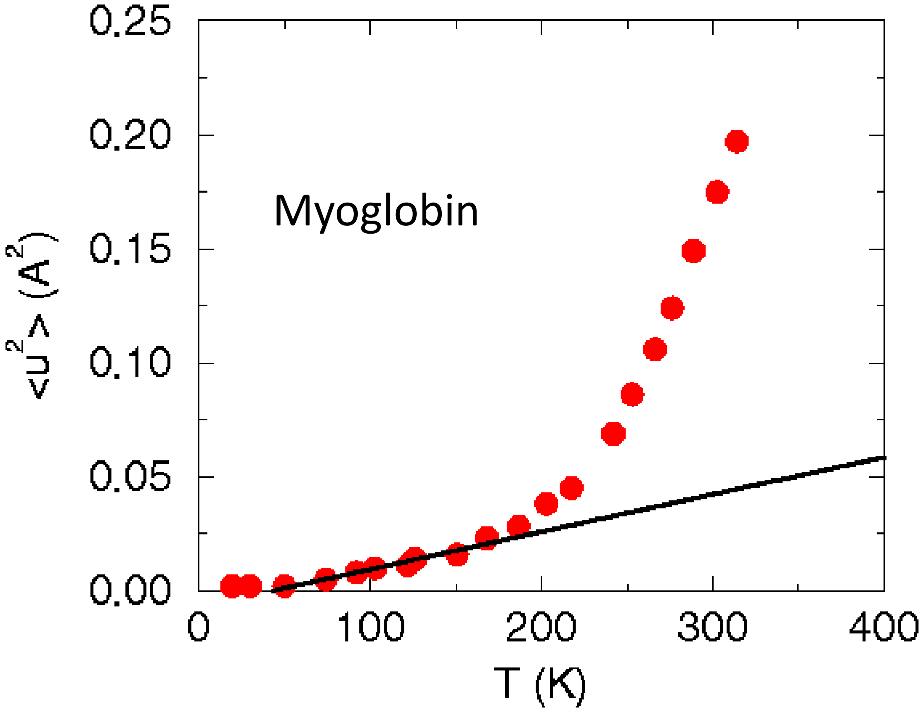

Methyl


CH_2 in ethyl chain

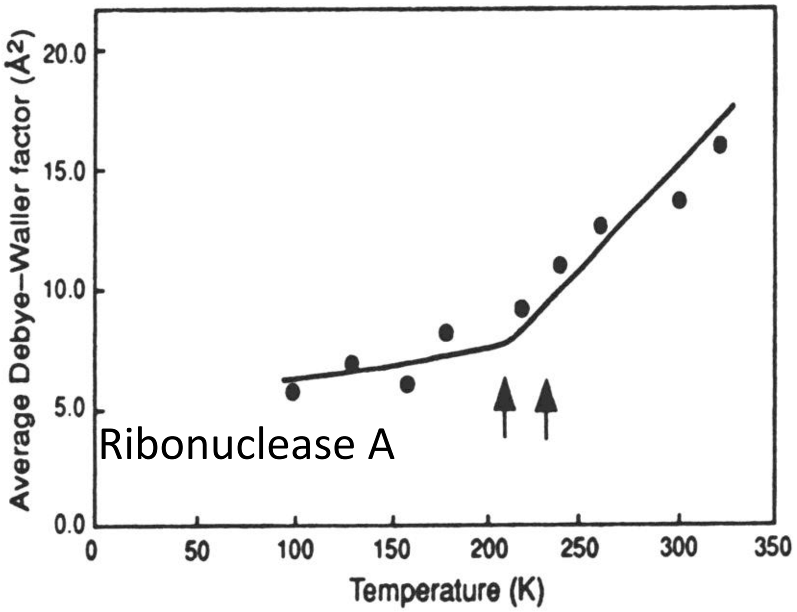

CH_3 in ethyl chain

Dynamic transition in proteins

From elastic neutron scattering experiments:



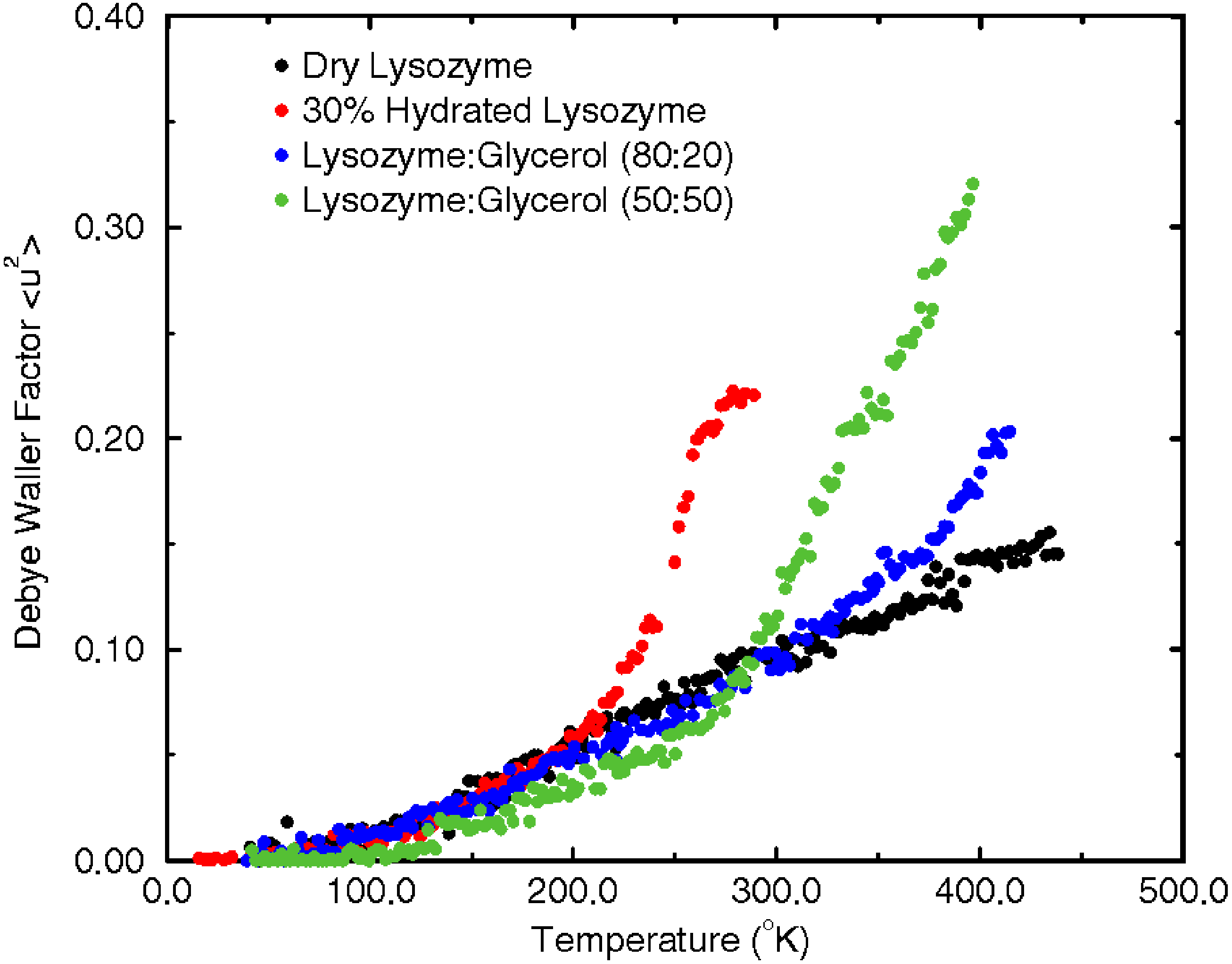
Doster *et al*, Nature (1989)



Rasmussen *et al*, Nature (1992).

Dynamic transition in $\langle u^2 \rangle$ correlates with the onset of enzymatic activity!

$\langle u^2 \rangle$ and proteins: Lysozyme in Glycerol

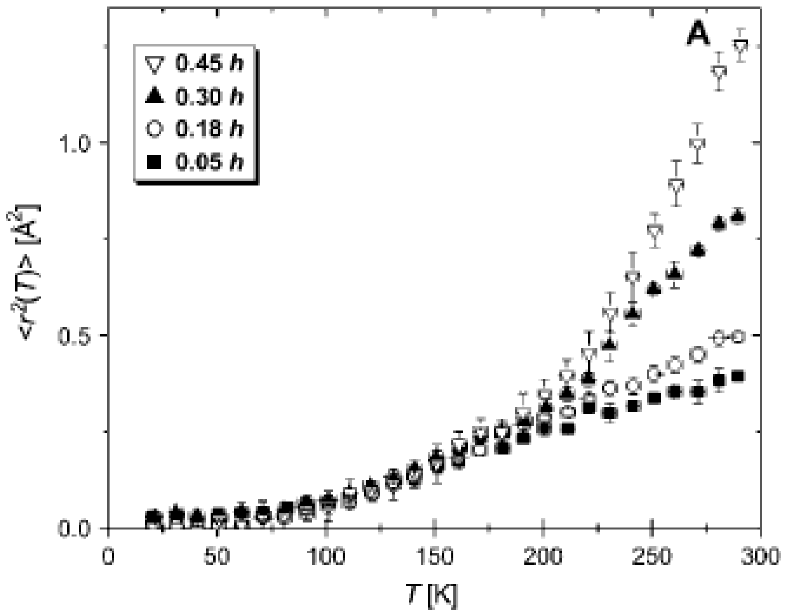


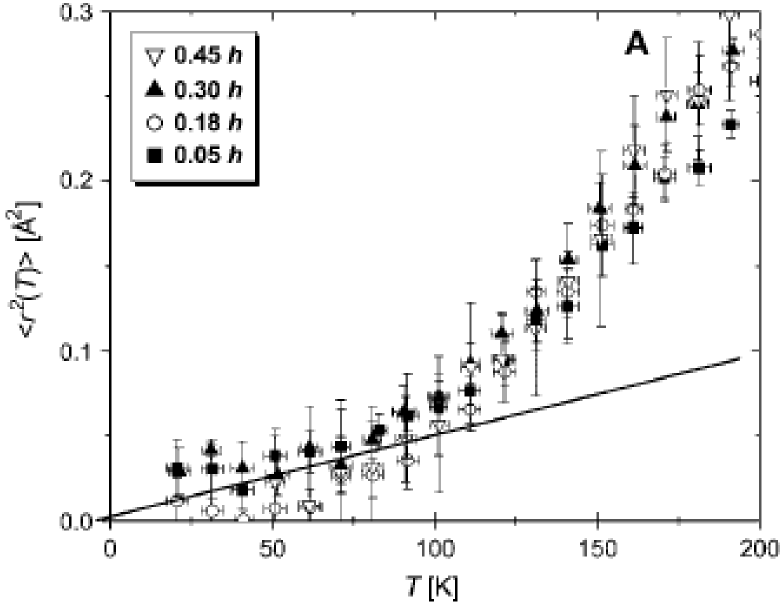
Tsai et al., Biophys. J. (2000)

Water must be present for the dynamic transition to occur

Glycerol retards the onset of anharmonicity

$\langle u^2 \rangle$ and proteins: Several transitions

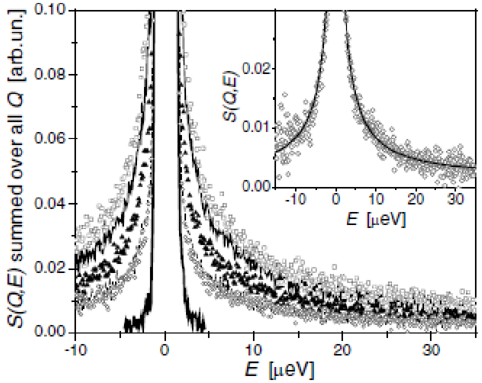




Roh et al., Biophys. J. (2006)

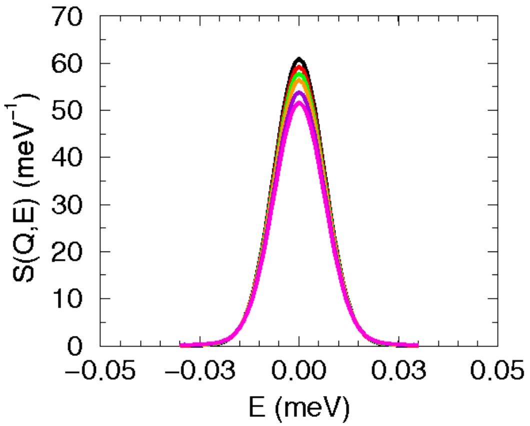
Lysozime:

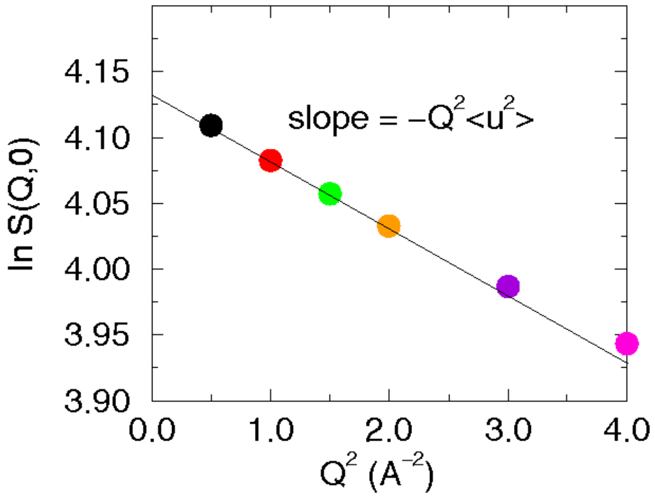
1. Standard dynamical transition at T_D~200-230 K.
2. Low-T onset of anharmonicity at T~100 K even for dry protein and related to the onset of methyl group rotation.

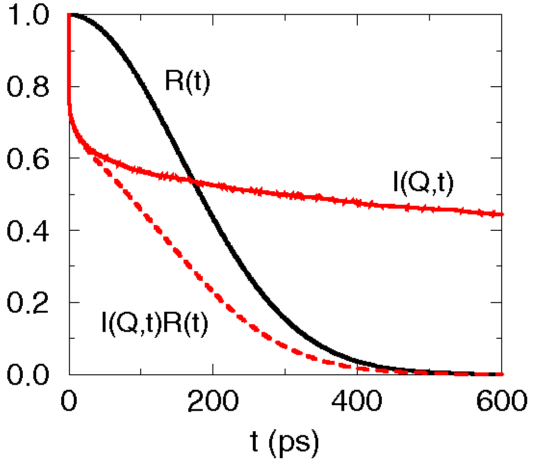


Calculation of $\langle u^2 \rangle$

Experiment: estimate mean-squared displacement from elastic intensity via Debye-Waller factor: $I(0) = \exp(-Q^2\langle u^2\rangle)$





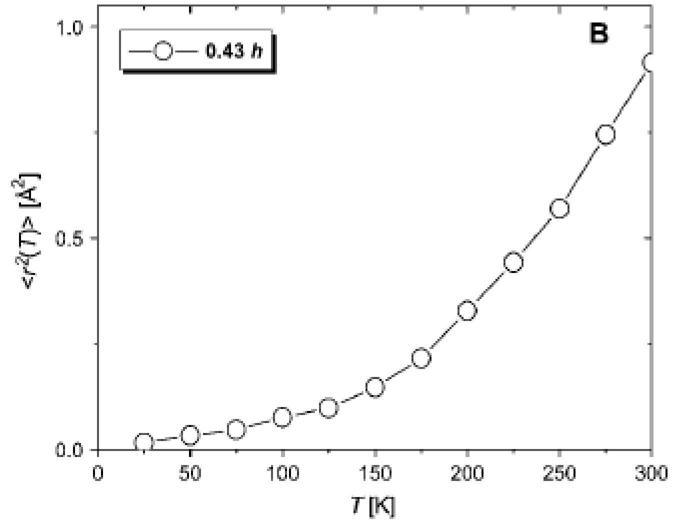


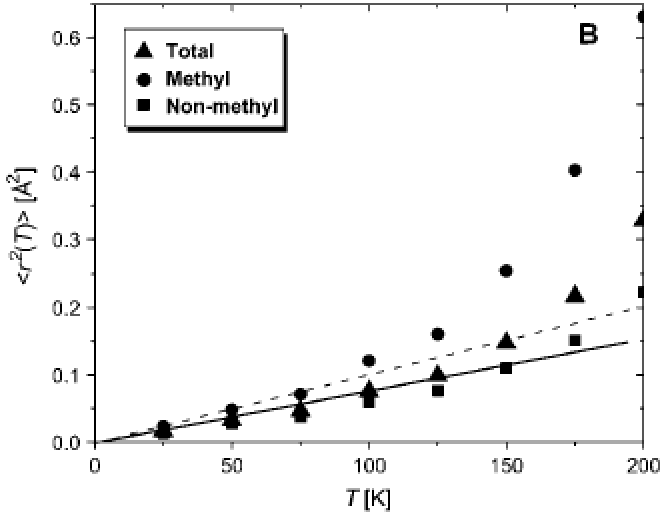
Simulation: calculate resolution-broadened $S(Q,E)$ as FT of $I(Q,t)R(t)$, where $R(t)$ is the FT of the instrument resolution function

$$I(Q,t) = \frac{1}{N}\sum_j e^{i\vec{Q}\cdot\vec{r}_j(t)} e^{i\vec{Q}\cdot\vec{r}_j(0)}$$

Methyl group dynamics and anharmonicity

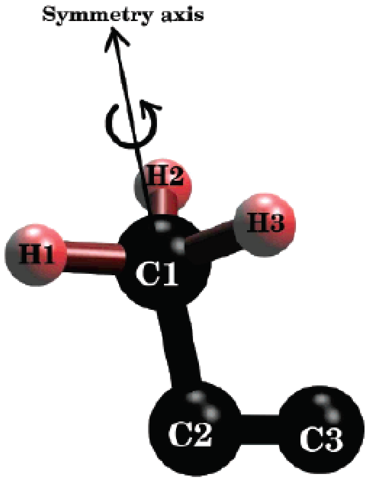
Lysozime

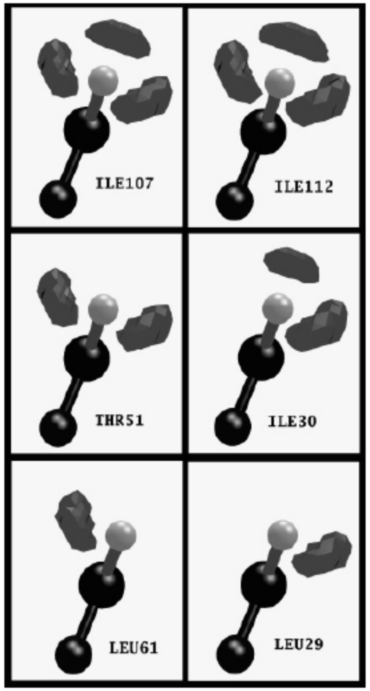


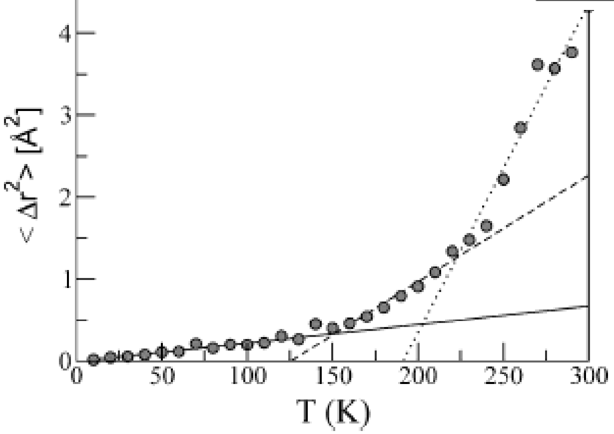


Roh et al., Biophys. J. (2006)

Crystalline myoglobin



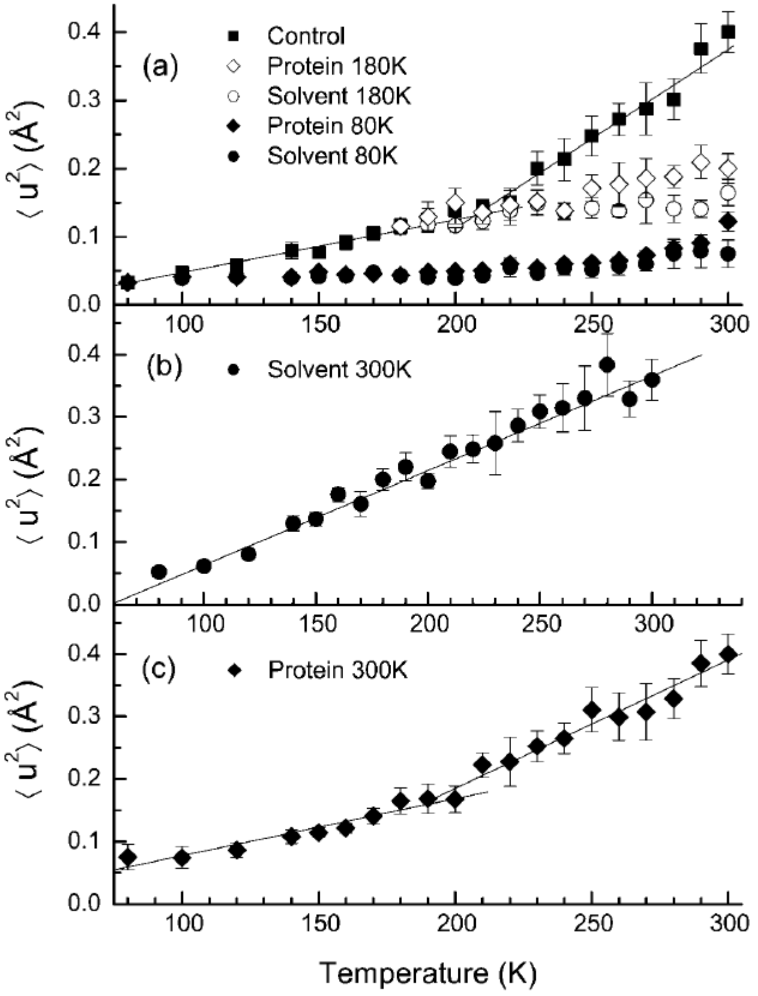


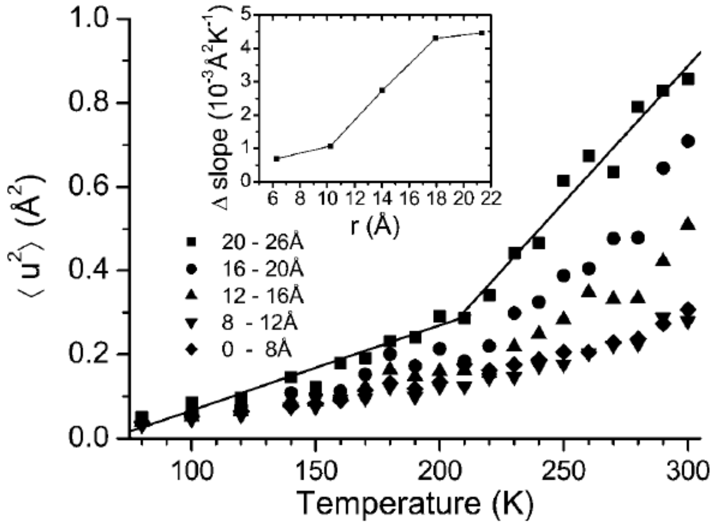


Krishnan et al., J. Phys. Chem. B (2008)

Protein-solvent coupling

Dual heatbath MD of 1 myglobin + 492 H_2O





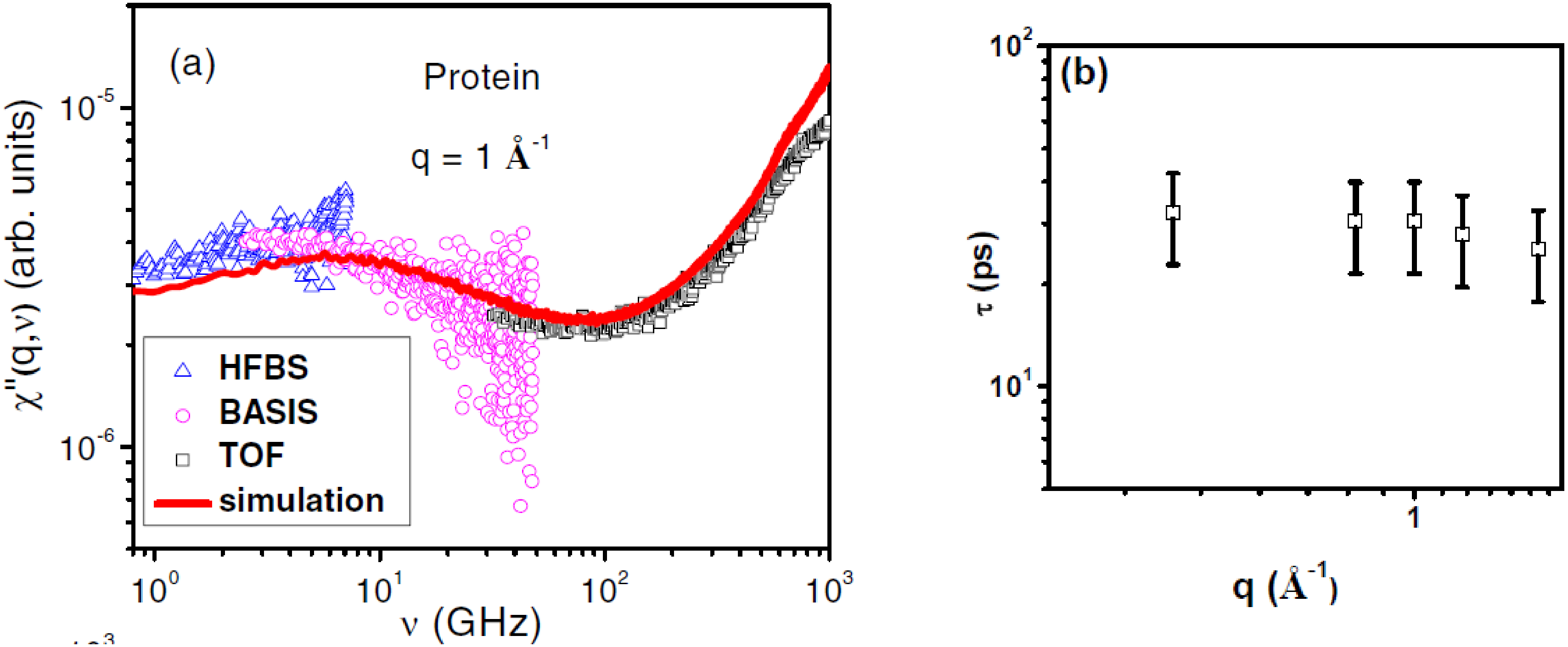
Low T solvent cages protein dynamics.

Dynamical transition more pronounced in outer parts of the protein.

Tournier et al., Biophys. J. (2003)

Motion decomposition in protein dynamics

Two randomly oriented lysozime molecules hydrated to h=0.3



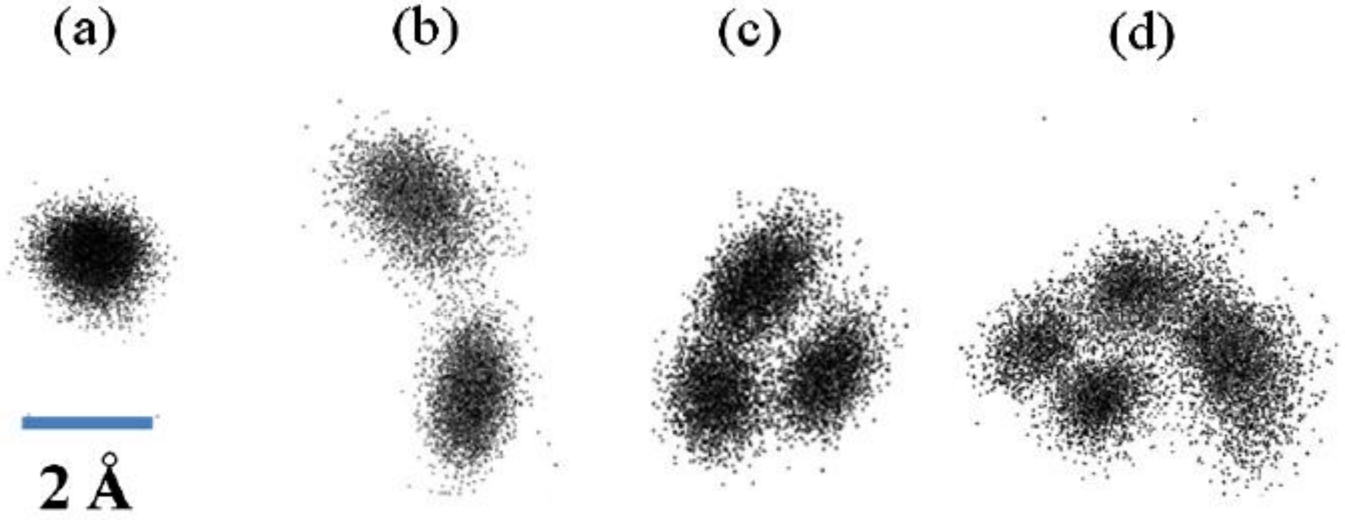
$$\chi''(Q,\omega)=S(Q,\omega)\times\left[\exp(\hbar\omega/kT)-1\right]^{-1}$$

Hong et al., PRL (2011)

Relaxation time independent of Q in the range 0.4 to 1.6 $Å^{-1}$ → Motion confined within a radius of ~3.5 Å.

Motion decomposition in protein dynamics

Coordinates of H atoms during 10 ns trajectory:



Hong et al., PRL (2011)

Two types of atoms:

I → Diffusing inside a single localized region (cluster)

II → Diffusing in clusters and occasionally jumping between them

+

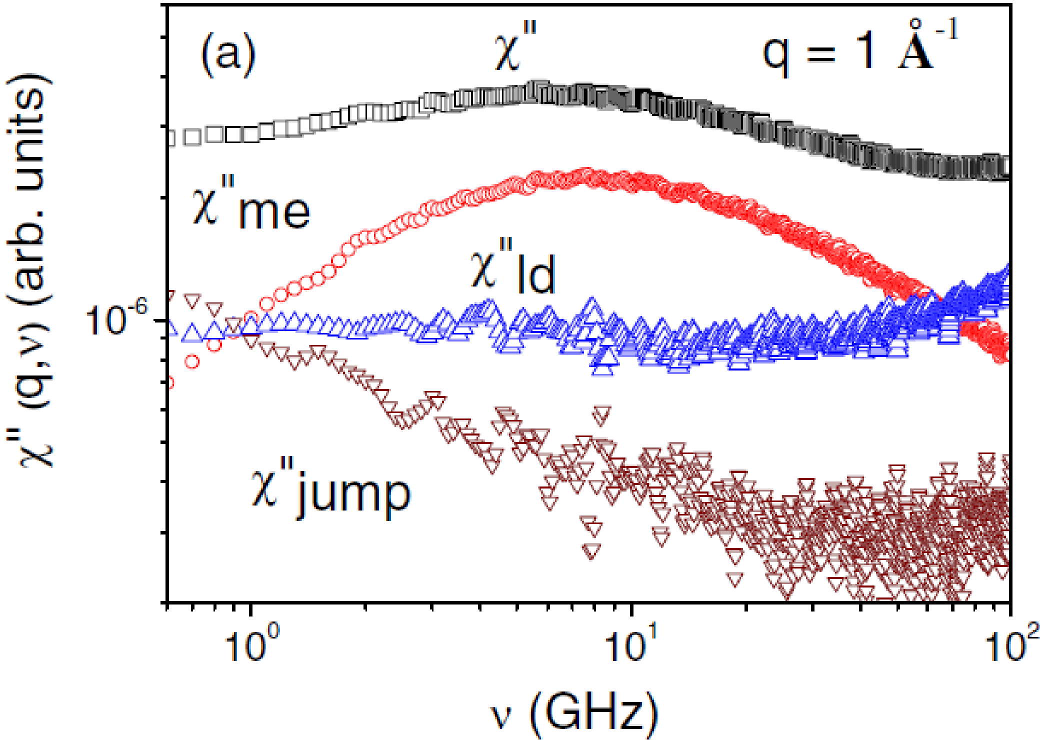
Methyl group rotations (3-fold rotations)

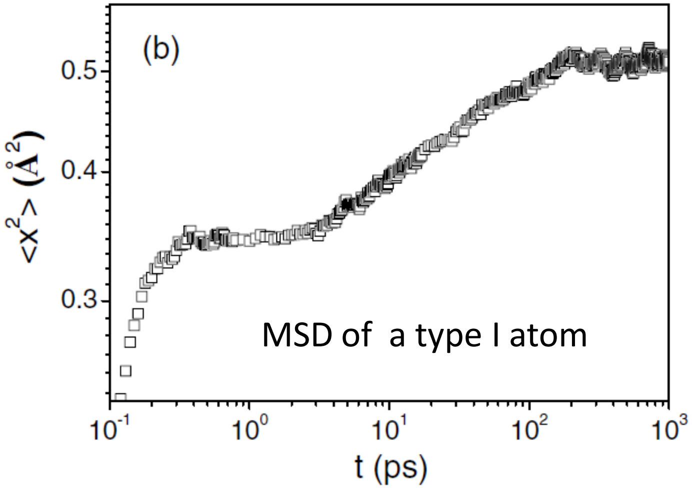
Motion decomposition in protein dynamics

Localized diffusion (type I) and broad cluster-size distribution.

Methyl group rotation with $\tau \approx 22$ ps.

Average relation time for jumps ~ ns.





MSD of a type I atom

Hong et al., PRL (2011)

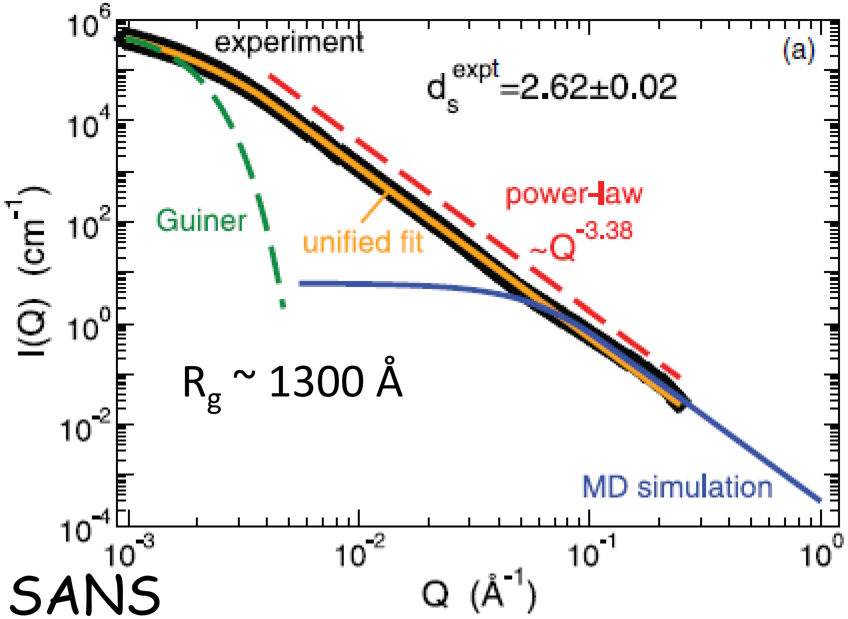
Model-free interpretation of the spectra in terms of a simple decomposition of H fluctuations in a globular protein.

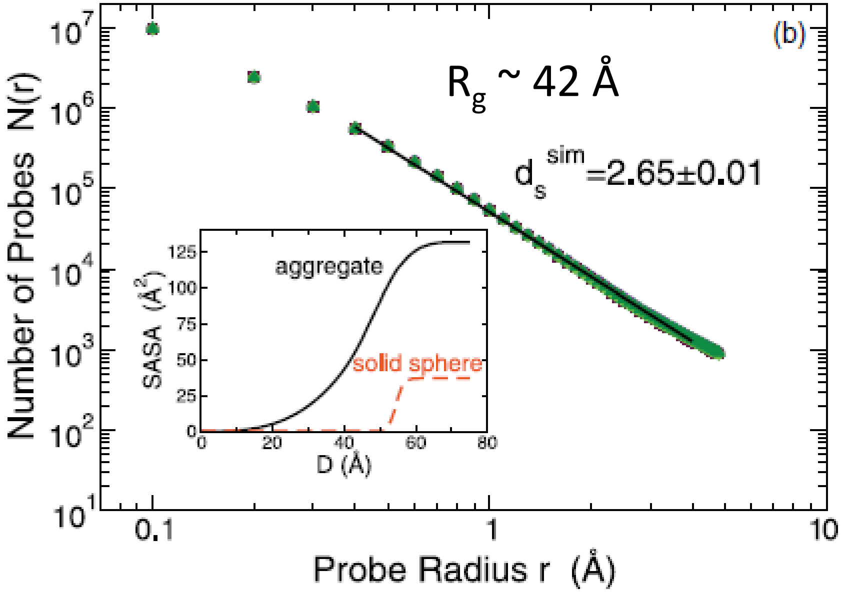
Multiscale structure in lignin

SANS: Structure on length scale ~10-1000 Å

MD: 25 lignin polymers (13 kDa) forming an aggregate of size ~84 Å (~400000 atoms → 10^6 processor hours ~ 100 years in single proc)

Petridis et al., Phys. Rev. E (2011)





SANS

$$I(Q) = G\exp(-Q^2R_g^2/3) + B(Q_*)^{-\alpha}$$

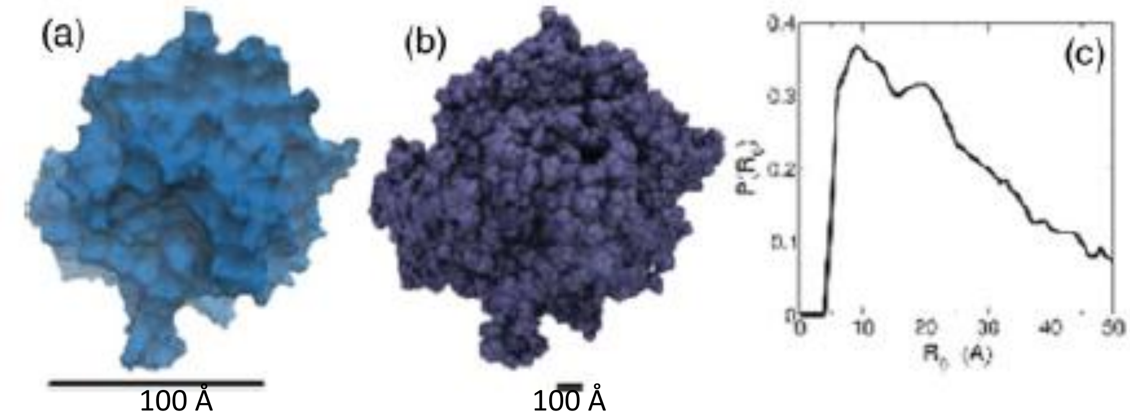
MD: $N(r) = \dfrac{A(r)}{2\pi r^2} \propto r^{d_s}$

A(r) = solvent accesible surface area (SASA)

Surface morphology invariant over length scales ~1-1000 Å

Multiscale structure in lignin

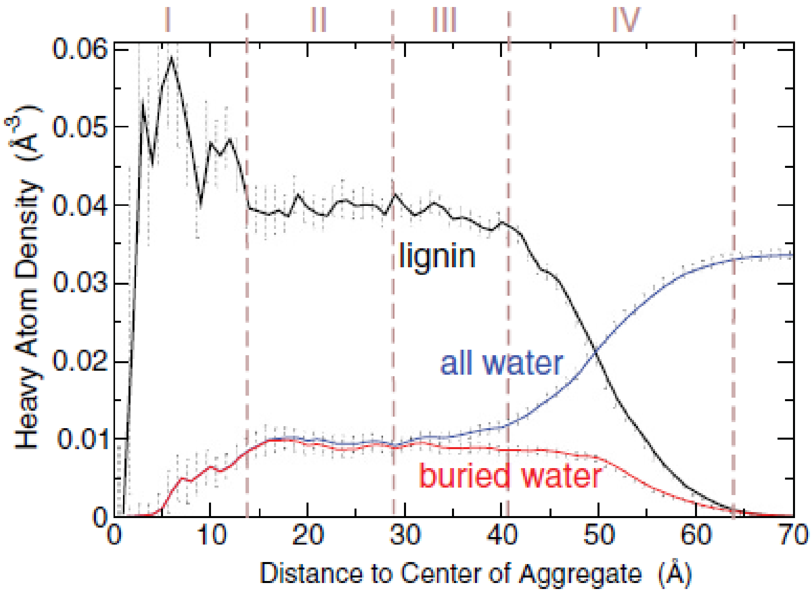
Used simulated small aggregates to construct larger ones:



Size distribution of surface pores → Many pores can bind to celulolytic enzymes (with radius of catalytic domains ~20 Å)

Lignin aggregates are hydrophobic, but they are significantly penetrated by water.

4 shells can be distinguished at different distances from the CM of the aggregate.



Petridis et al., Phys. Rev. E (2011)

Summary

- Simulations are a very effective tool to obtain a better understanding of complex systems.

- They allow to investigate features and properties that are not easily accessible experimentally.

- Today simulations are easily accessible to experimentalists. Good programs and resources are freely available and are relatively simple to use.

Software and references

Codes:

- DFT : VASP, CASTEP, Siesta, Abinit, CPMD, Materials Studio...
- FF-MD : NAMD, DL_POLY, Gromacs, LAMMPS, Tinker, Gulp, Materials Studio ...

Books:

- Allen & Tildesley: *"Computer simulation of liquids"*, Oxford University Press (1987)
- Frenkel & Smit: *"Understanding Molecular Simulation"*, Academic Press (1996)
- Richard M. Martin: "*Electronic Structure*", Cambridge University Press (2004)