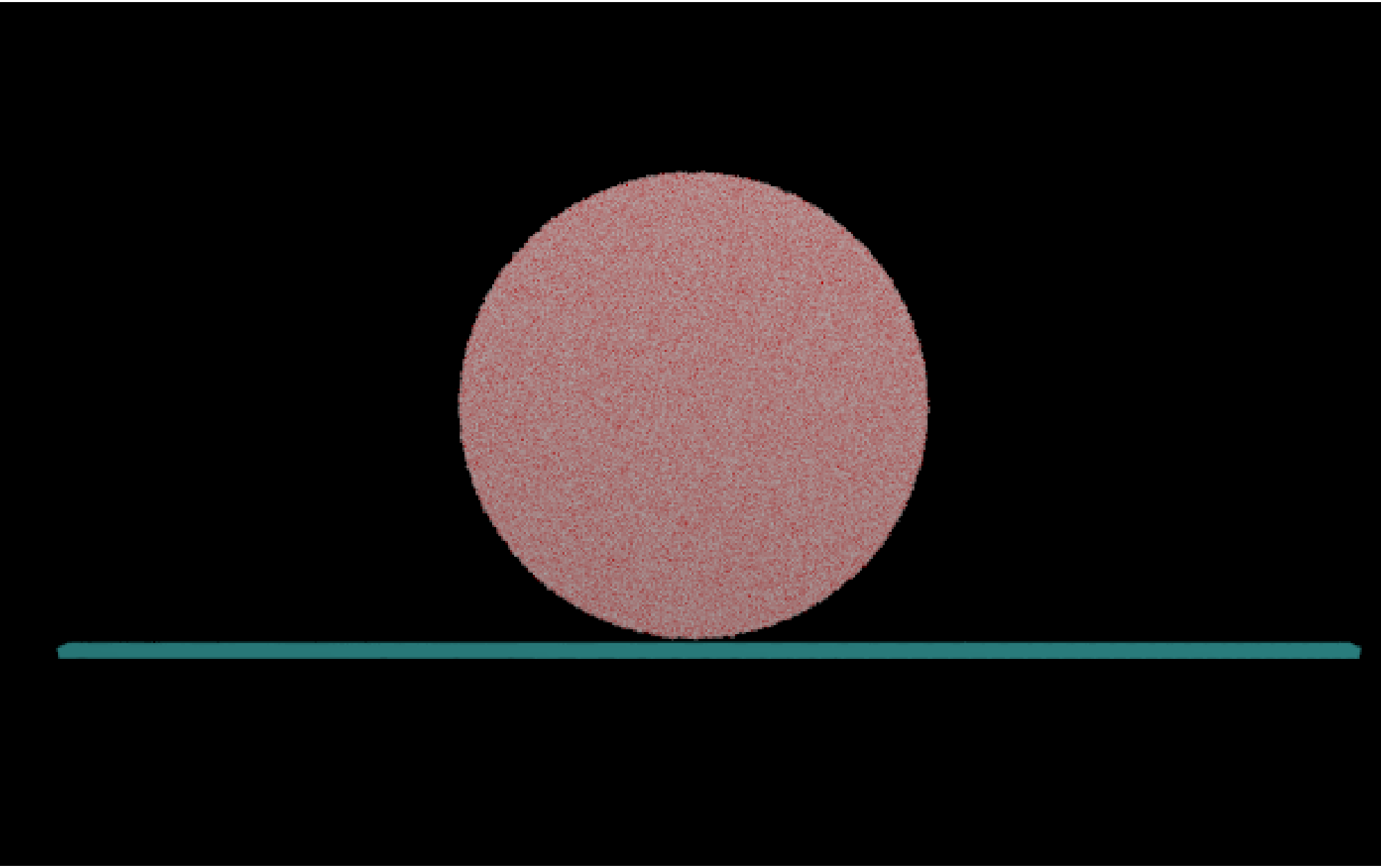
Introduction to molecuclar dynamics

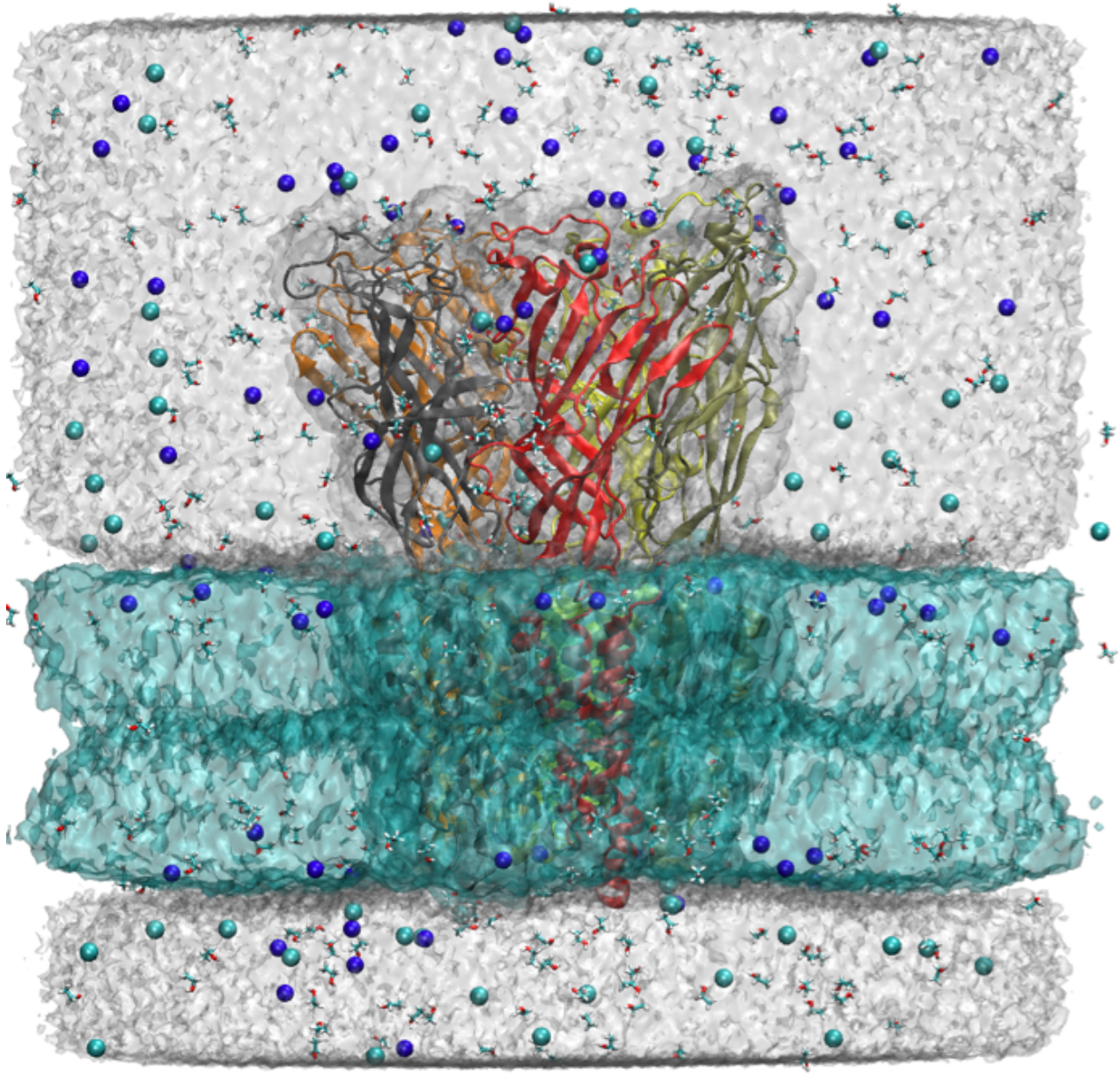
Mark Abraham, BioExcel and KTH, Stockholm

GROMACS Development Manager

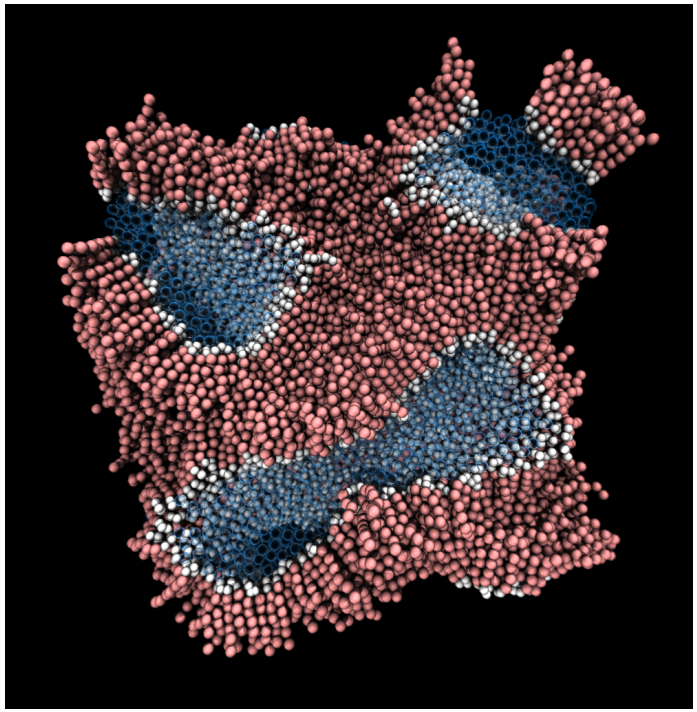
Molecular dynamics - flow of droplets

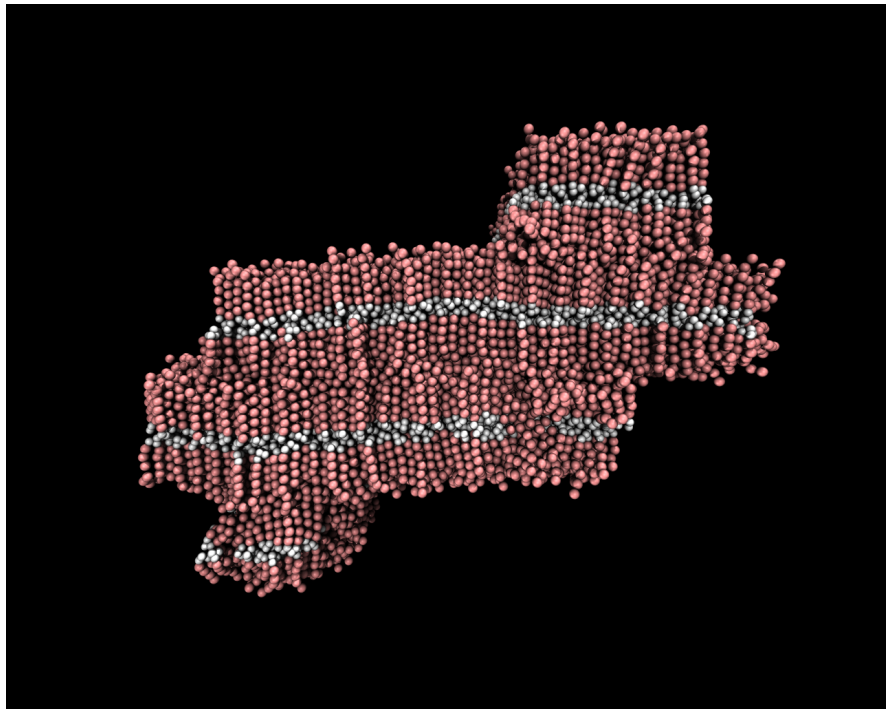


Molecular dynamics - membrane proteins



Molecular dynamics - skin formation



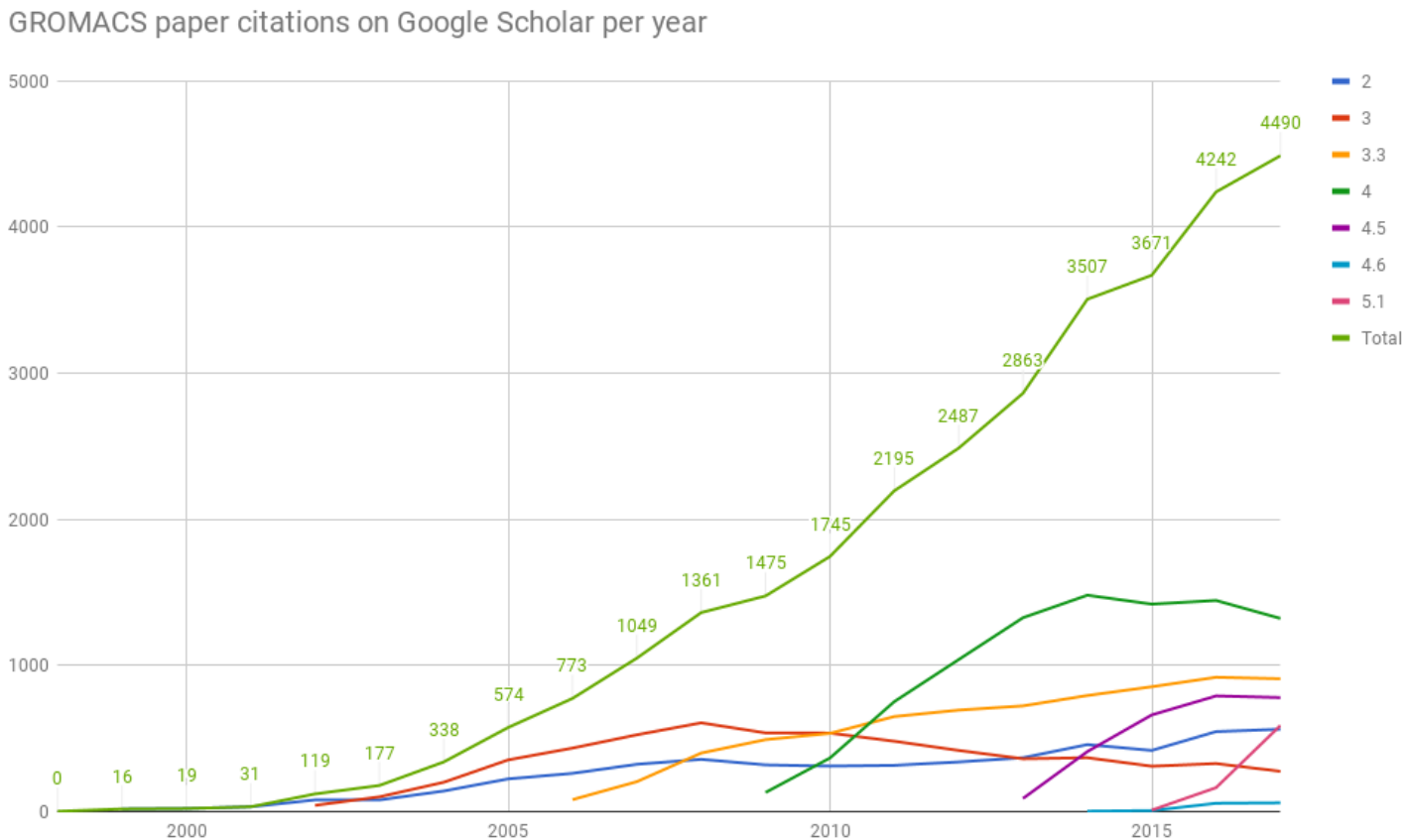


GROMACS

- classical molecular dynamics
- mostly targets problems from biochemistry
- widest range of force fields
- wide range of simulation algorithms

GROMACS

- Free and open-source C++11 community project
- Developed across multiple institutions
- Used by hundreds of research groups

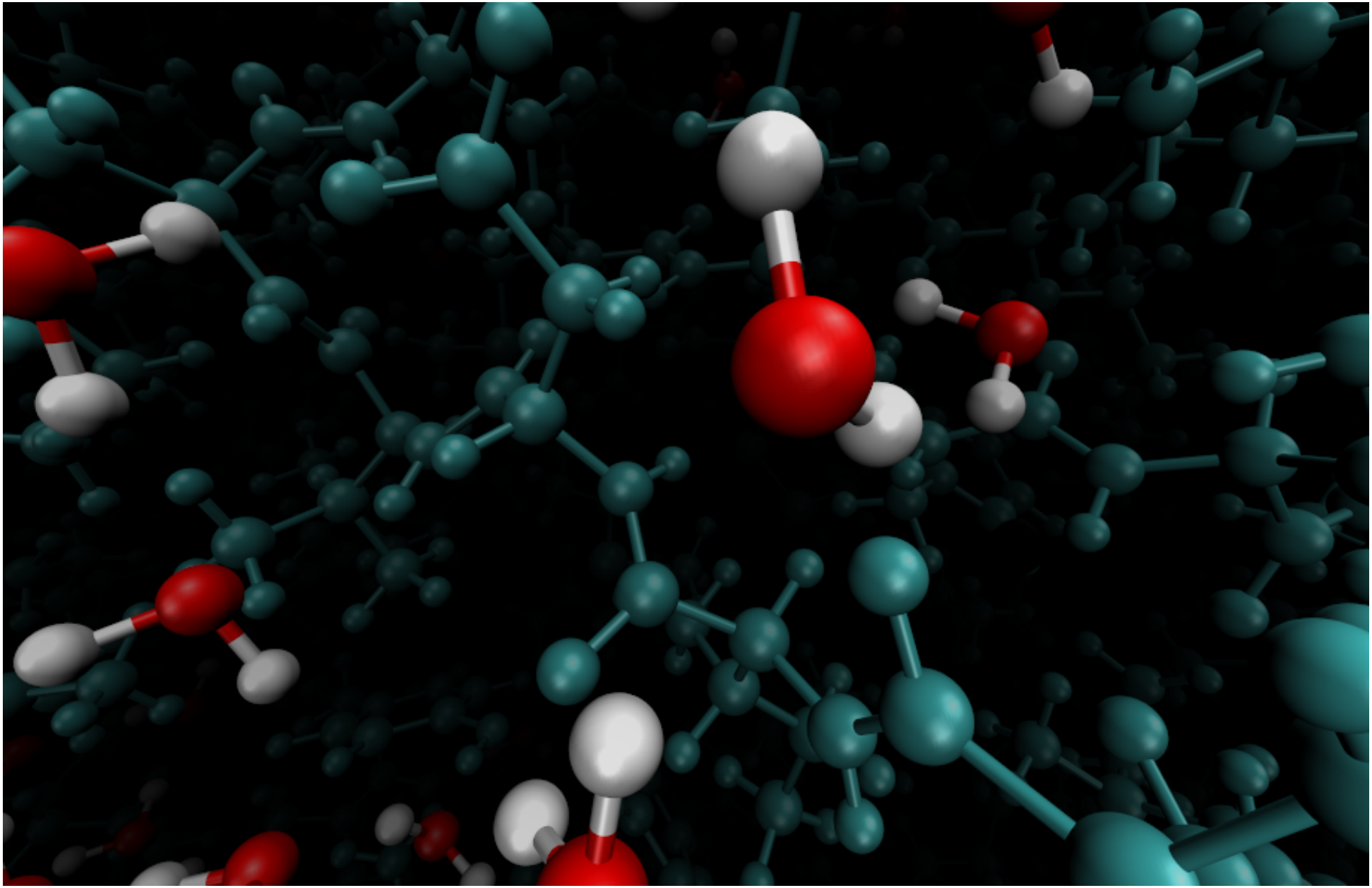


Molecular dynamics

$$\mathbf{m}\frac{\mathrm{d}^2\mathbf{x}}{\mathrm{d}t^2} = -\frac{\partial V}{\partial \mathbf{x}}$$

- Iterative solver for a set of second-order differential equations.
- Given initial conditions $\mathbf{x}$ and $\mathbf{v}$, compute forces.
- Use the forces to propagate the system over a finite timestep.
- Repeat. Lots.
- Hope that the sampling of the conformational ensemble is ergodic.
- Then the time averages match the ensemble averages observed in the real system

Molecular mechanics force fields



Molecular mechanics force fields

$$V = \sum_{i \in \text{bonds}} \frac{a_i}{2} (d_i - d_{i0})^2 + \sum_{i \in \text{angles}} \frac{b_i}{2} (\theta_i - \theta_{i0})^2$$
$$+ \sum_{i \in \text{torsions}} \frac{V_n}{2} (1 + \cos(n\omega_i - \gamma_n))$$
$$+ \sum_{\text{all}, i<j} V_{ij}(\mathbf{r}_{ij})$$

That final term covers all of the non-bonded interactions, typically Coulomb and van der Waals treated seperately.

Parameters for MM force fields

- determined empirically, based on experimental and QM results
- usually based on atom types, often many for each element
- intimately co-dependent on functional forms and each other
- many families of related force fields (AMBER, CHARMM, GROMOS, OPLS, Martini, . . .)
- do not expect to mix and match
- can be extended to treat new kinds of molecules (here be dragons)
- use online servers / offline tools where possible (SwissParam, ATB, Antechamber, acpype, PRODRG, STaGE, MKTOP)

How much work is a biomolecular MD simulation?

One typically needs

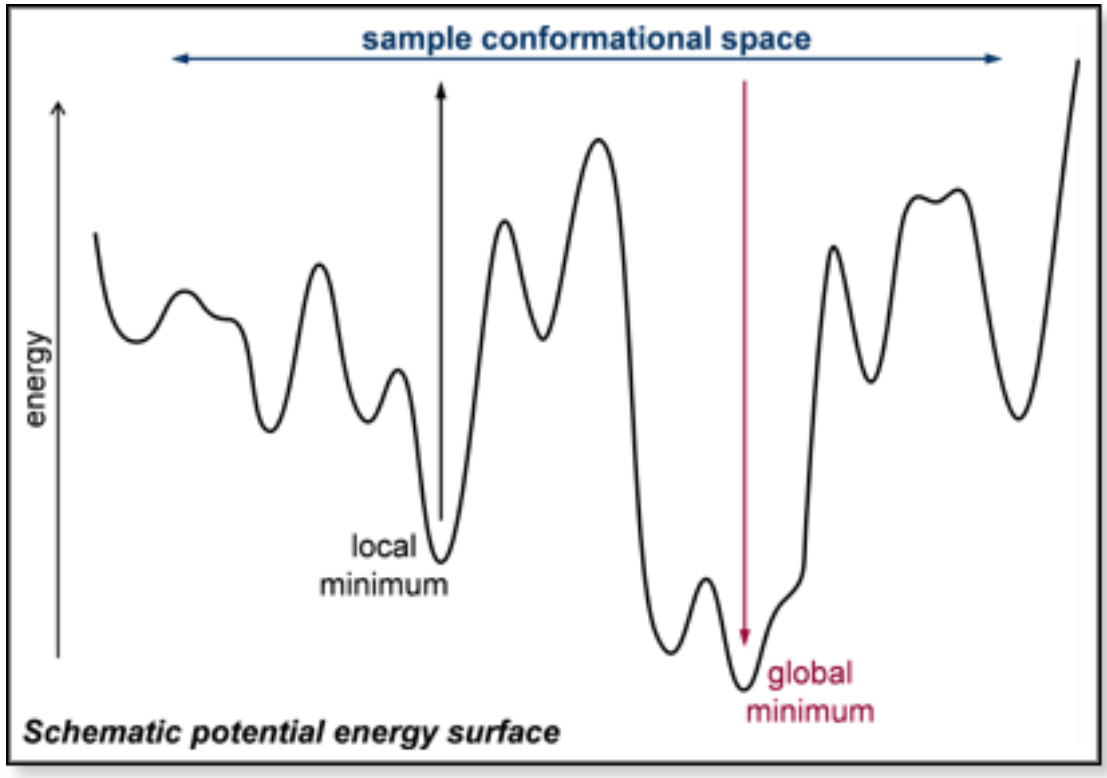
- time steps around 2 femtoseconds (2×10^{-15} s),
- for at least a million steps (to reach nanosecond time scales),
- for something like 150,000 particles,
- where each particle has around 800 interactions, and
- each interaction costs around 100 flops,
- (plus a bit of other stuff)

which comes to

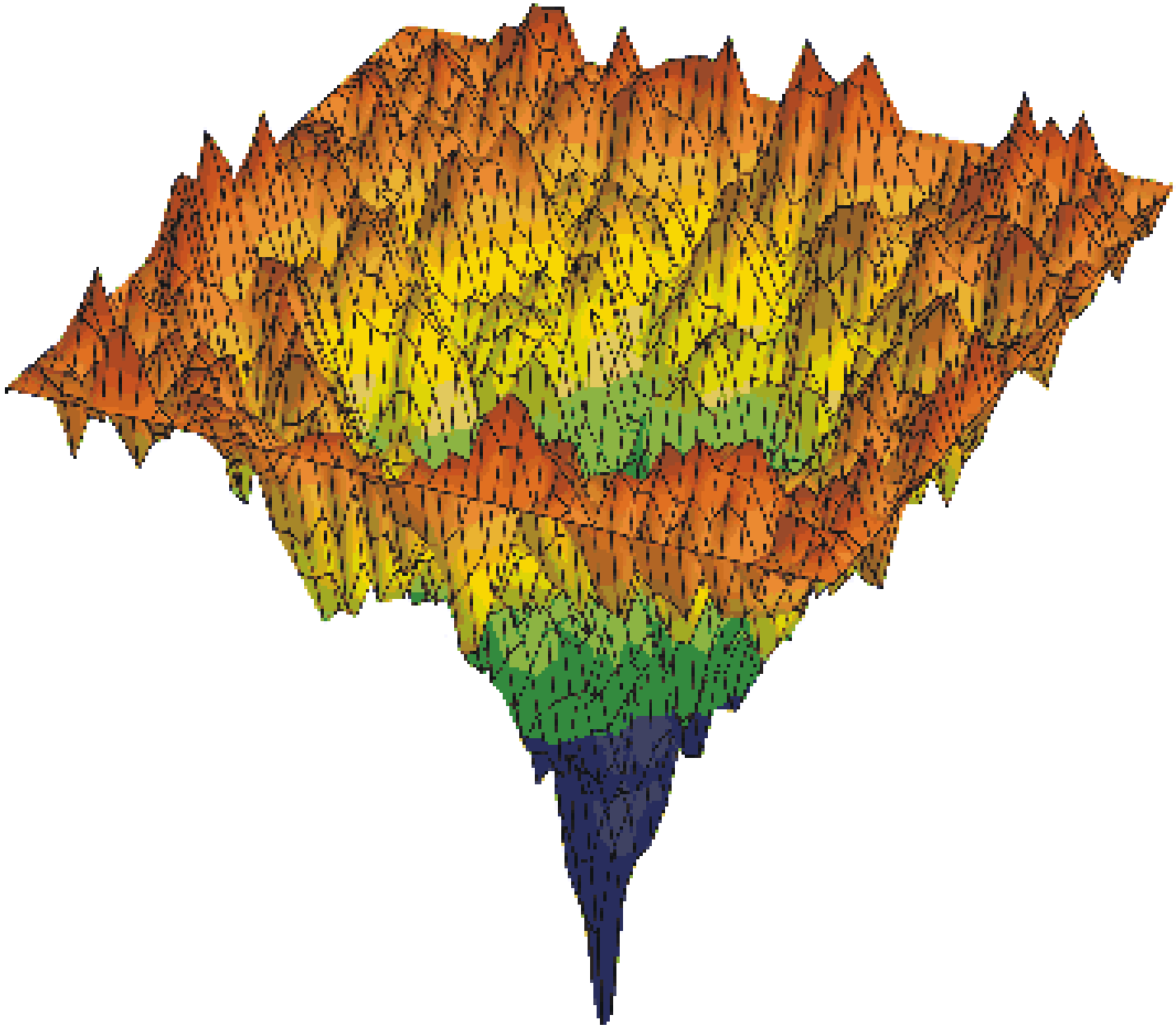
$$10^6 \times 1.5 \times 10^5 \times 200 \times 100 = 1 \times 10^{16}$$

flops

That's several days on a current-generation workstation, so clearly we need high quality fast implementations, and to work smarter rather than harder.

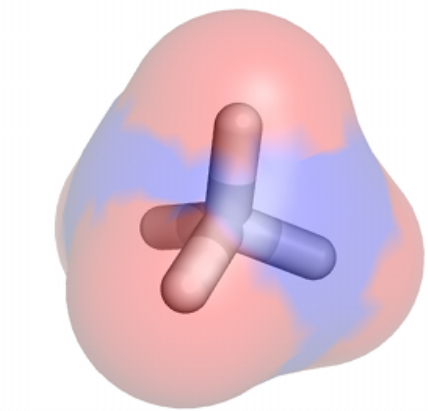
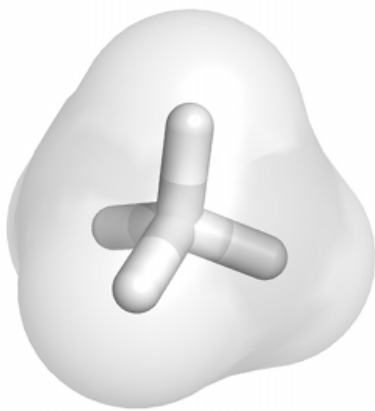
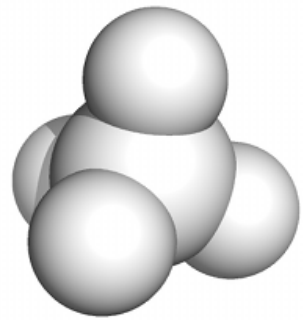

sample conformational space
energy
local
minimum
global
minimum
Schematic potential energy surface

High-dimensional energy surfaces

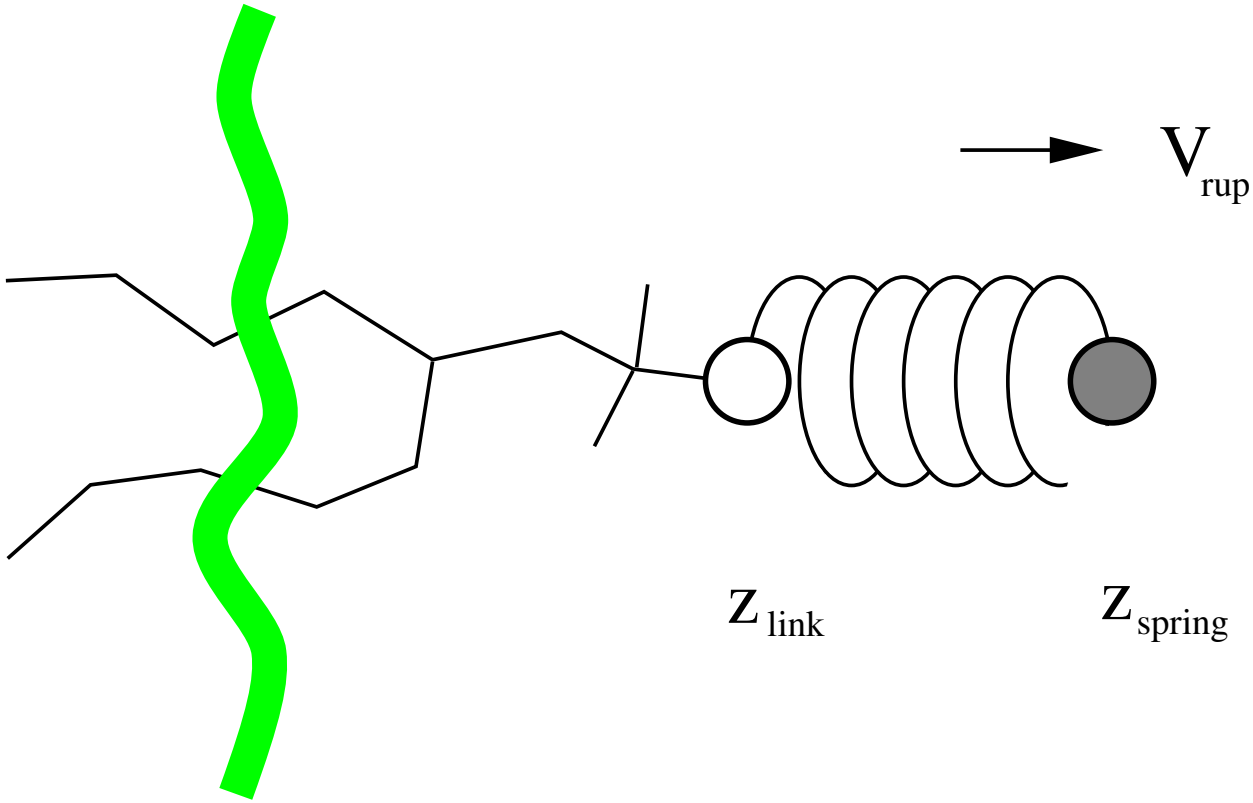


Free-energy calculations

- Often it's too hard to sample the motion that you want
- Thermodynamics doesn't care about paths
- Many different kinds of techniques to compute free energy differences

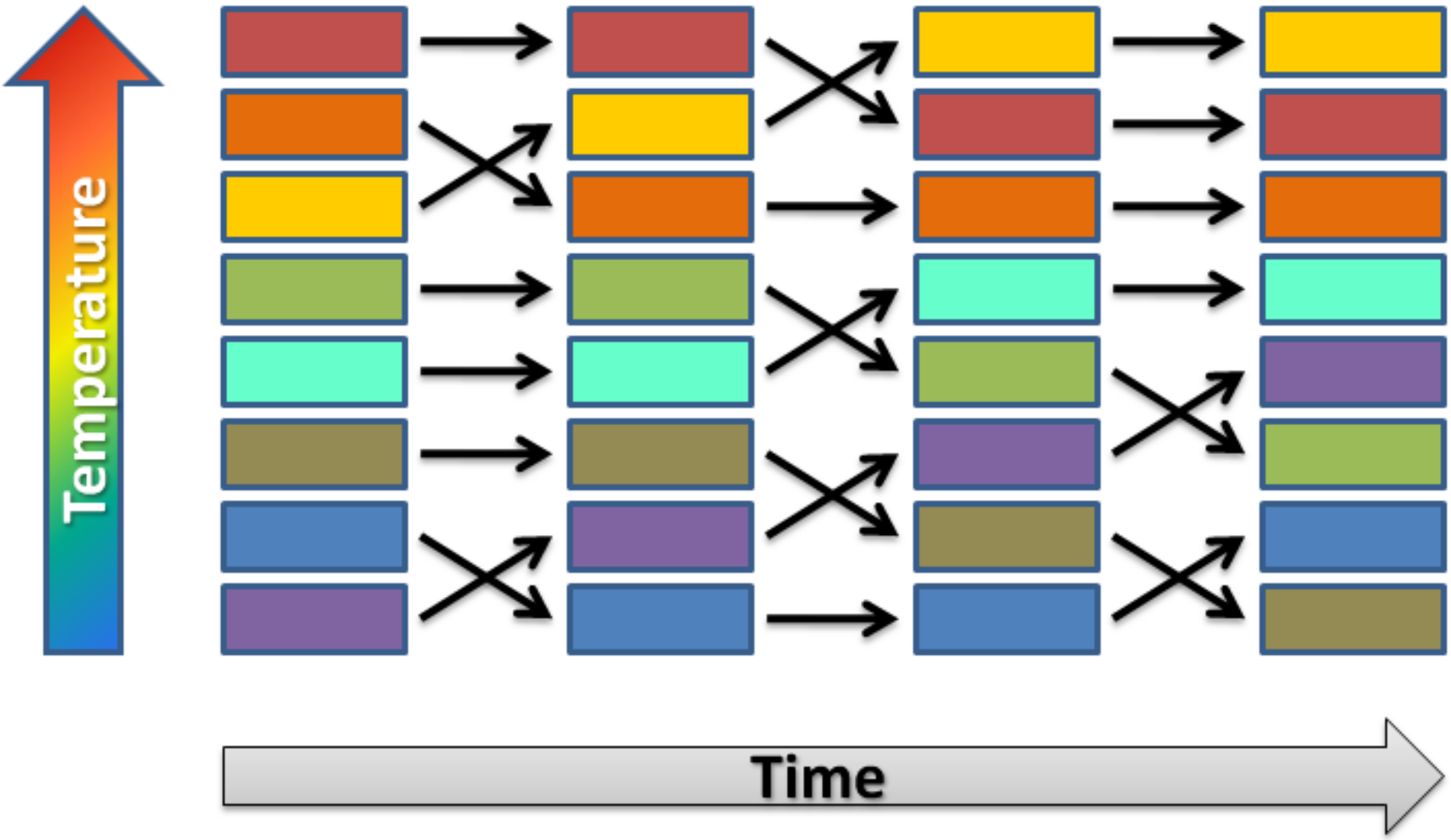




Computing potentials of mean force



- umbrella sampling
- AWH

Replica exchange



Other features in GROMACS

- Many integrators, coupling types
- Constraints, virtual sites
- Analysis tool suite (selection syntax)
- Restrained simulations
- User-supplied tabulated interactions (bonded and non-bonded)
- Expanded ensemble simulations
- Lennard-Jones PME
- Computational Electrophysiology
- Essential dynamics
- Normal-mode analysis
- Simulated annealing
- Interactive Molecular Dynamics with VMD
- GPU support, including some auto-tuning
- Polarizable models
- PLUMED metadynamics
- Third-party setup tools

So you want to do molecular dynamics?

Do you *really* want to do MD?

You should avoid modelling

- processes making and breaking chemical bonds
- systems with mobile charges (metals, CNT, TM complexes)
- phase transitions
- systems with strong polarization effects

You should prefer

- systems with regular repeating units
- systems similar to those previously modelled
- systems with well defined states (e.g. ionizable groups)

Choosing a validation target

If you just want to do a calculation, fire away, you'll get numbers and have no idea what they're good for.

But if you want to do a simulation of the real world that has predictive value, you need to plan how you will show that.

- your software, force field, method parameters are all known to work
- you have tested that they are working e.g. reproduce past work
- you will be able to sample enough to have statistical validity

Choosing a model resolution

For classical molecular simulations, the major choices are

- atomistic (all-atom or united-atom; CHARMM, AMBER, OPLS, GROMOS)
- coarse grained (Martini)

Basic trade-off between the amount of detail available from the model, versus the amount of work required to compute the details for that more complex model.

Remember that these models were designed to work with particular simulation packages and methods within them... avoid moving outside this region!

Choosing software for scientific modelling

Often, you should use the tool that you can get local help with.

Seek training opportunities with other tools and methods.

If you only know how to use a hammer, you'd better hope you don't come across a screw!

Look for published studies comparing accuracy, performance, and feature sets of comparable tools.

Choosing the model contents

- pH?
- membrane composition
- what multimer? are they really the same?
- co-solvents, ligands, ions (and relative placement of these)

How will you express these in your molecules' topologies?

Choosing an initial configuration

MD is dynamical, so it moves stochastically from one point in the target ensemble to another. Where you start *should* be irrelevant!

Often some experimental structure, possibly after homology modelling is the starting point, but you should remember that the process of determining a structure can mean that that structure is different from ones found in the simulation environment. Modelling is tough!

Choose a simulation cell that is large enough, but not too large!

Forming an equilibration protocol

We always need to carefully relax our initial configuration into something from which we can simulate

- starting structures often have defects
- experimental structures come from different ensembles
- initial velocity generation is random

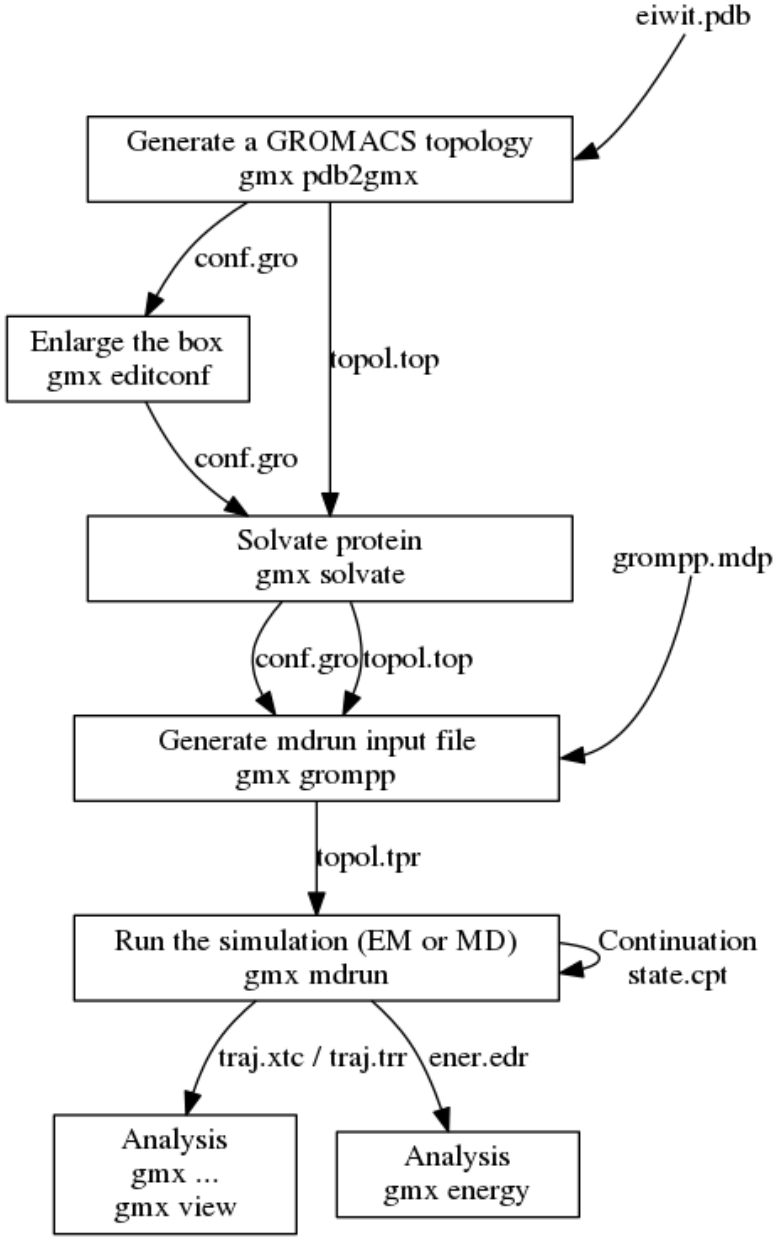
Be prepared to use multiple stages, with the early ones very, very gentle.

Choosing a simulation ensemble

Equilibration aims to reach a particular target ensemble

- conceptually simplest ensemble has constant energy
- biology tends to take place at constant temperature and pressure
- don't just copy the input from the last project!
- follow established methods others have used and published

System preparation with GROMACS tools



GROMACS resources

http://manual.gromacs.org/documentation/

gmx-users mailing list -
http://www.gromacs.org/Support/Mailing_Lists

The end - any questions?