CHEM 498Q / 630Q

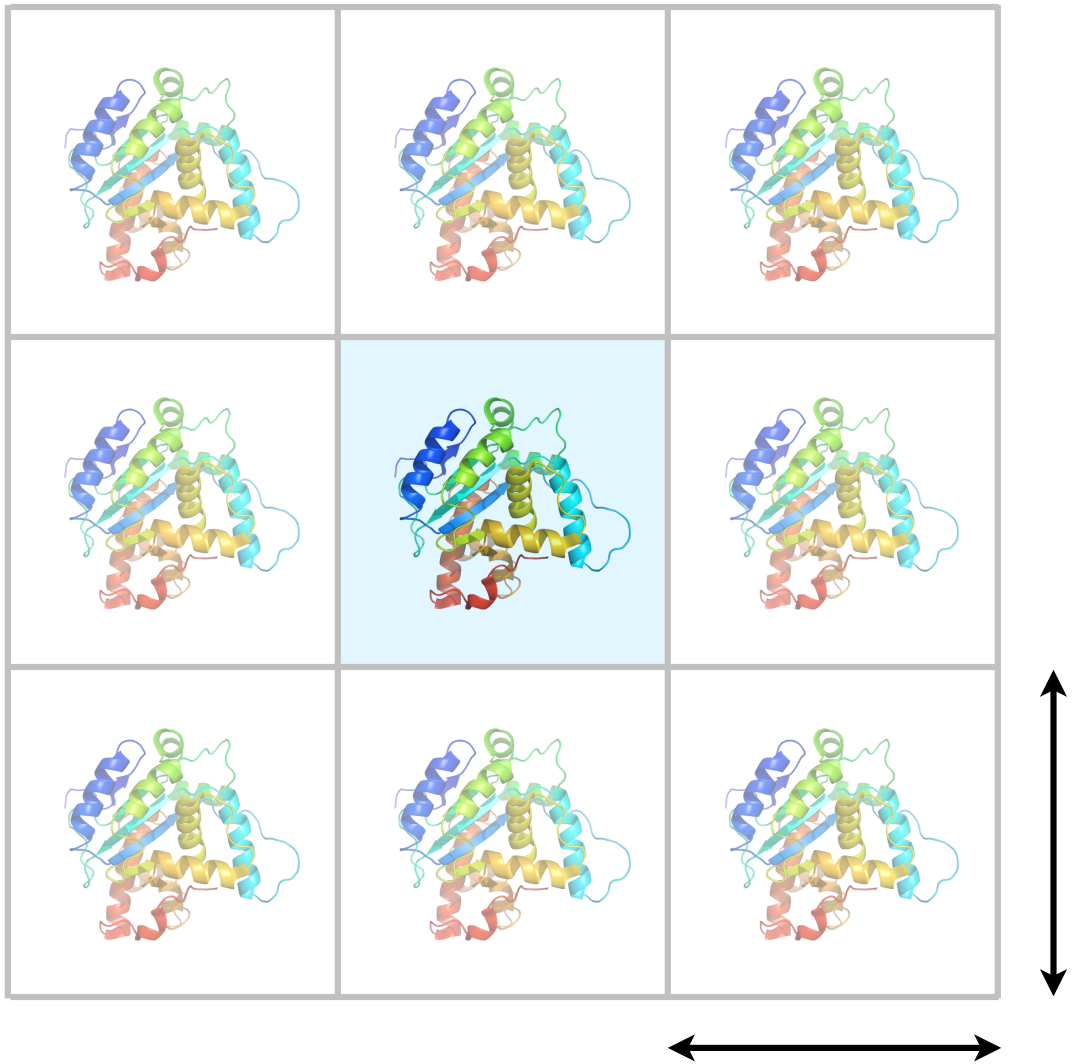
Molecular modelling of proteins

Fall 2015 Term

Instructor:
Guillaume Lamoureux
Concordia University, Montréal, Canada

Periodic boundary conditions

(with electrostatic energy calculated using an Ewald summation)



Lattice size should be large enough to accommodate tumbling motions of the solute and to preserve a layer of solvent of **at least 2 Debye lengths** between the images.

Numerical integration

$$m_i \mathbf{a}_i(t) = \mathbf{F}_i(t)$$

Non-integrable system of differential equations

Finite-difference approximations :

$$\mathbf{a}_i(t) \simeq \frac{\mathbf{v}_i(t+\frac{1}{2}\delta t) - \mathbf{v}_i(t-\frac{1}{2}\delta t)}{\delta t}$$

$$\mathbf{v}_i(t+\tfrac{1}{2}\delta t) \simeq \frac{\mathbf{r}_i(t+\delta t) - \mathbf{r}_i(t)}{\delta t}$$

$$\mathbf{v}_i(t-\tfrac{1}{2}\delta t) \simeq \frac{\mathbf{r}_i(t) - \mathbf{r}_i(t-\delta t)}{\delta t}$$

Finite-difference version of Newton's equations :

$$m_i \frac{\mathbf{r}_i(t+\delta t) - 2\mathbf{r}_i(t) + \mathbf{r}_i(t-\delta t)}{\delta t^2} = \mathbf{F}_i(t)$$

We can calculate $\mathbf{r}_i(t+\delta t)$ from $\mathbf{r}_i(t)$ and $\mathbf{r}_i(t-\delta t)$:

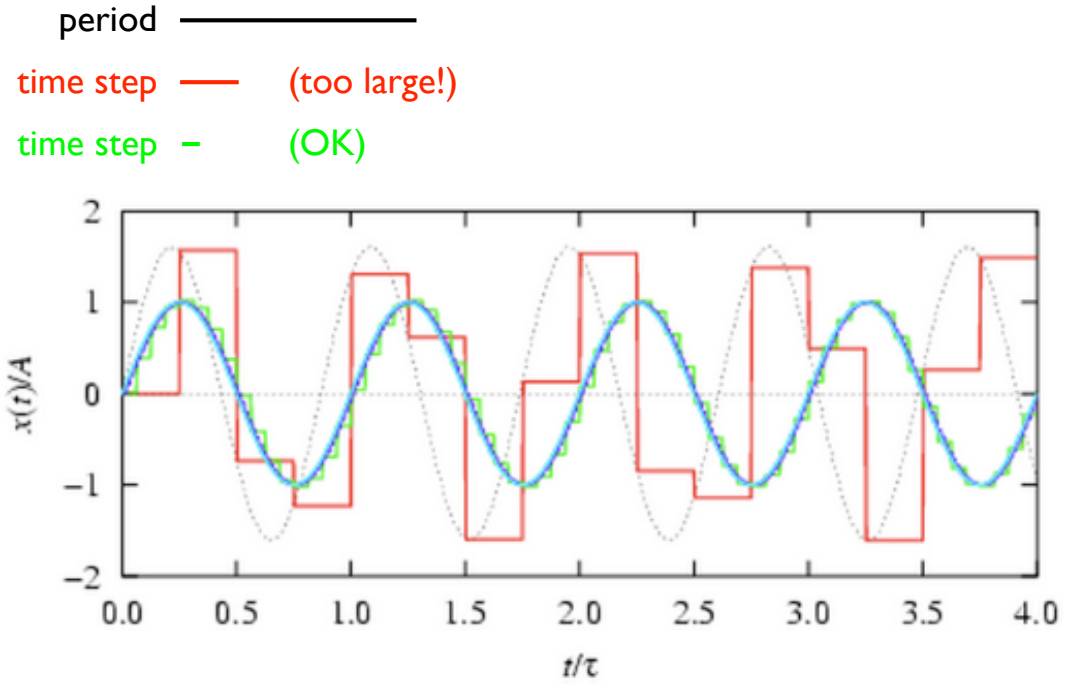
$$\mathbf{r}_i(t+\delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t-\delta t) + \delta t^2 \frac{\mathbf{F}_i(t)}{m_i}$$

(Verlet formula, 1967)

Choice of time step (δt)

Rule of thumb :

δt should be at least one tenth of the shortest period of oscillation of the system, but ideally one twentieth.



O–H stretch : $\sim 3500\ \text{cm}^{-1} \Rightarrow \tau \sim 10\ \text{fs} \Rightarrow \boldsymbol{\delta t \sim 0.5\ \text{fs}}$

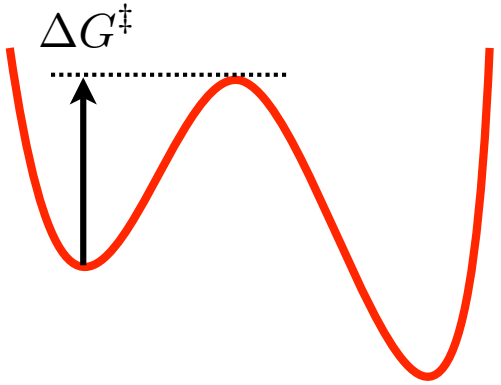
C=O stretch : $\sim 1700\ \text{cm}^{-1}$
N–O stretch : $\sim 1600\ \text{cm}^{-1}$
C=C stretch : $\sim 1600\ \text{cm}^{-1} \quad \Rightarrow \boldsymbol{\delta t \sim 1.0\ \text{fs}}$

To avoid using a 0.5 fs time step, the X–H bonds can be kept rigid (using the SHAKE algorithm).

How long should the simulation be ?

Typical time for crossing an energy barrier :

$$\tau = \tau_o \, e^{\Delta G^{\ddagger}/k_{\mathrm{B}}T}$$



Using $\tau_0 = 1$ ps :

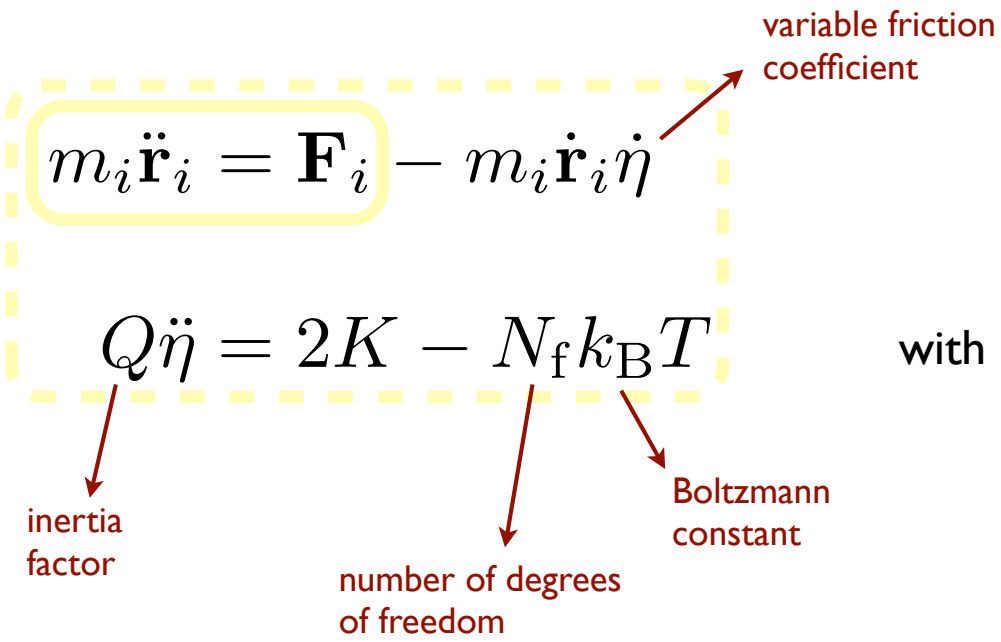
1 kcal mol^{-1} $\Rightarrow$ **1.2 ps^{-1}** ("1.2 crossings per ps")

5 kcal mol^{-1} $\Rightarrow$ **1.4 ns^{-1}**

10 kcal mol^{-1} $\Rightarrow$ **< 1 ms^{-1}**

The problem gets worse if the system has multiple barriers…

"Thermostatted" Newton's equations (Nosé–Hoover equations)



with $2K = \sum_j m_j \dot{r}_j^2$

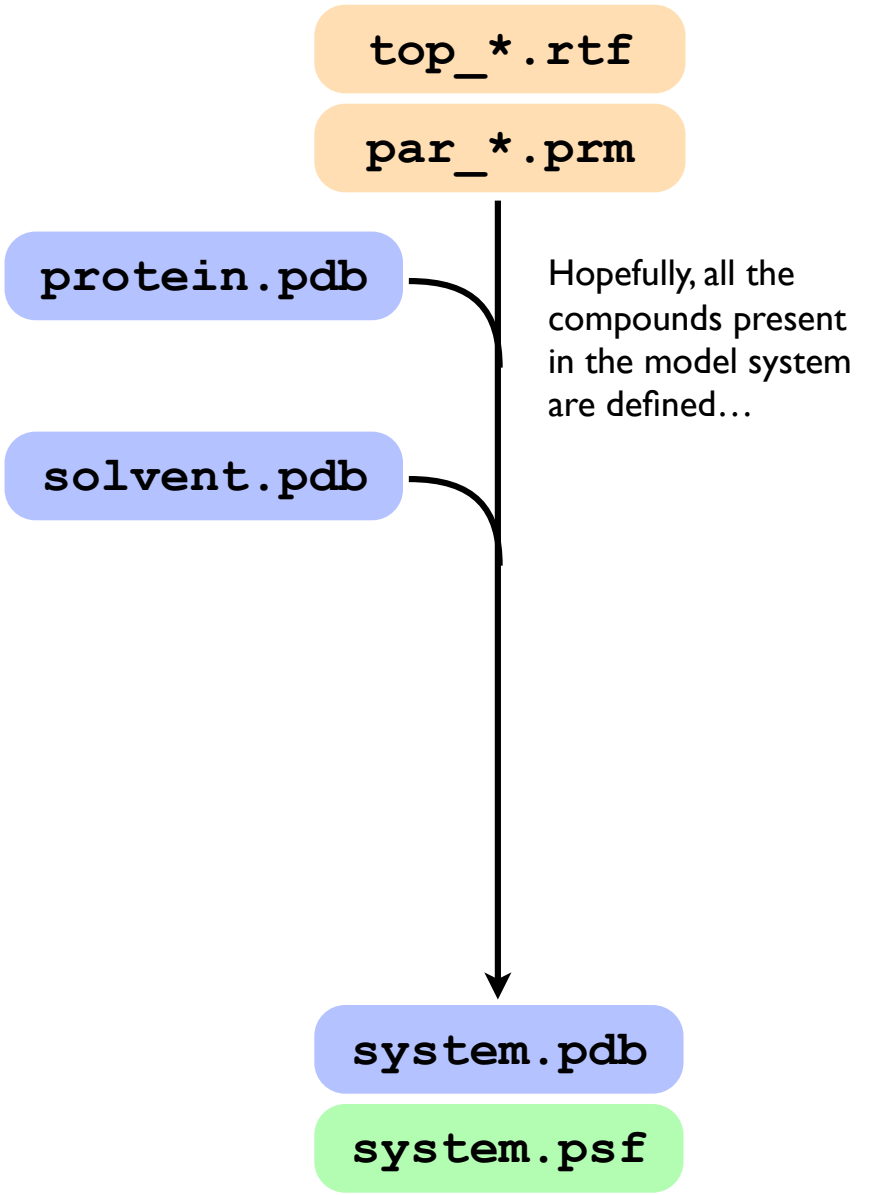
2 times the total instantaneous kinetic energy of the system

$N_\mathrm{f} = 3N - N_\mathrm{c}$

According to the equipartition theorem, each degree of freedom of the system should have a kinetic energy of $\frac{1}{2}k_\mathrm{B}T$.

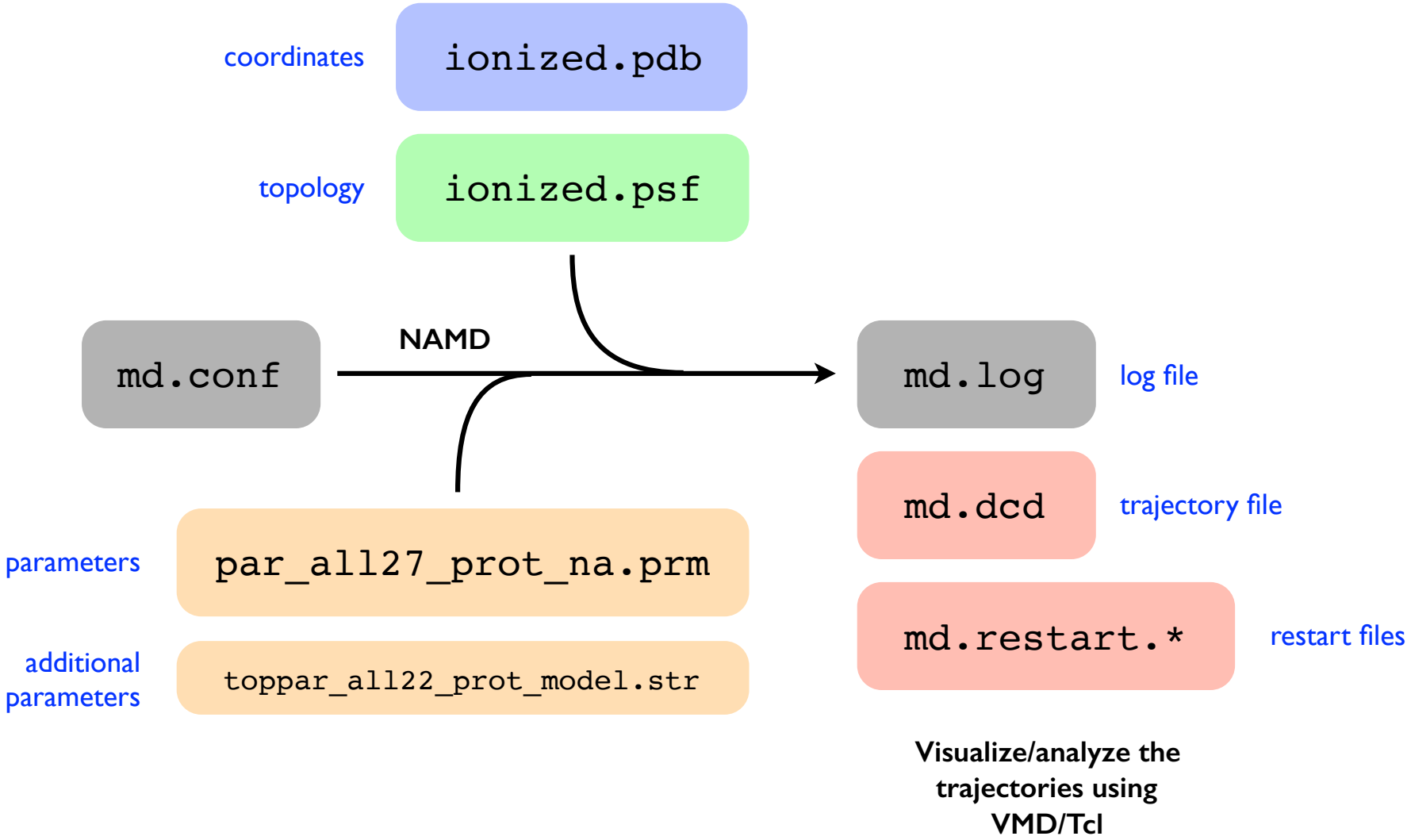
For that reason, *K* is related to the instantaneous temperature of the system.

General procedure for system setup

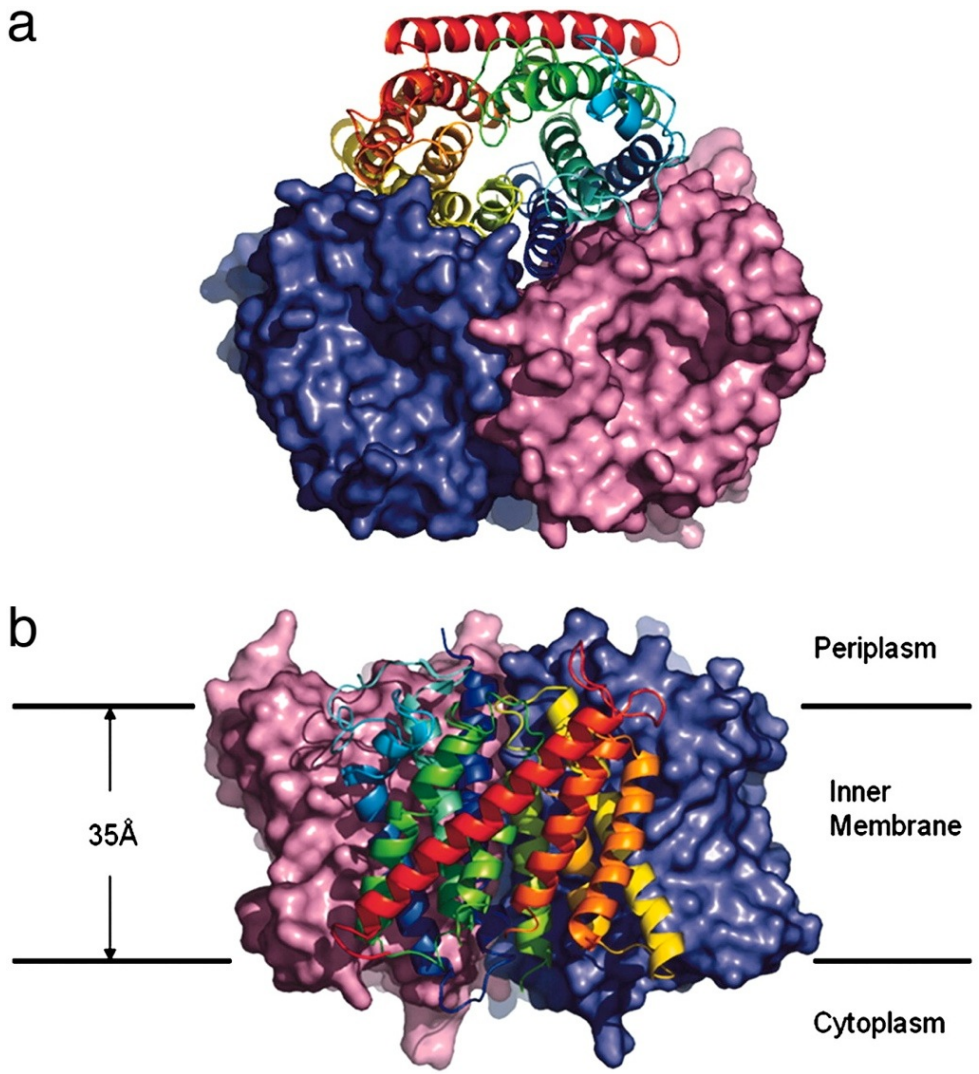


- **Read topology and parameter files**
- **Read PDB file**
  (protein coordinates)
- **Add solvent / membrane**
- **Specify details of energy function**
  (cutoff scheme, periodic boundary conditions, additional restraints, etc.)
- **Minimize energy**
  (to remove potential bad contacts)
- **Write PDB and PSF**
  (coordinates and topology)

MD simulation in NAMD

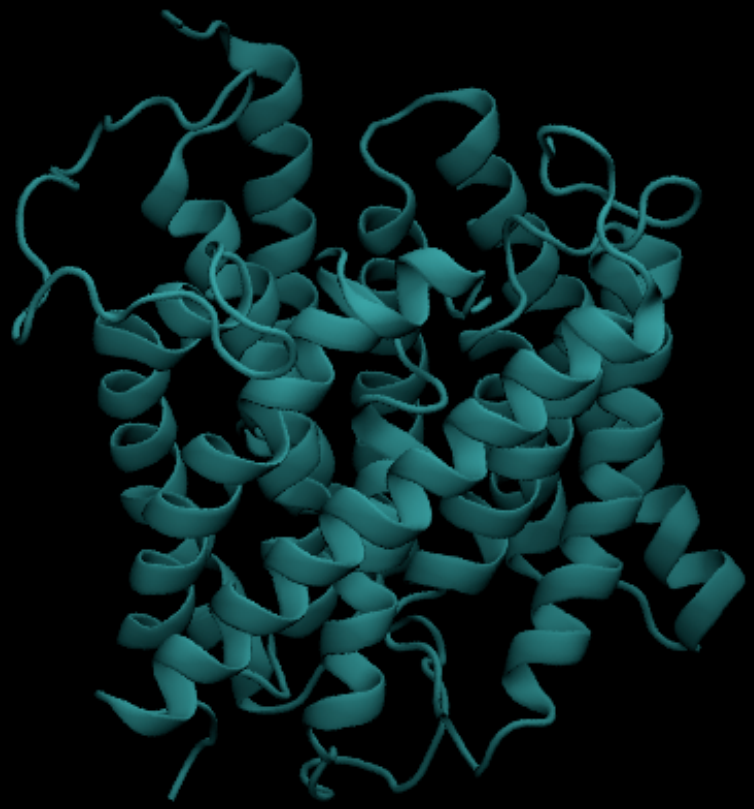


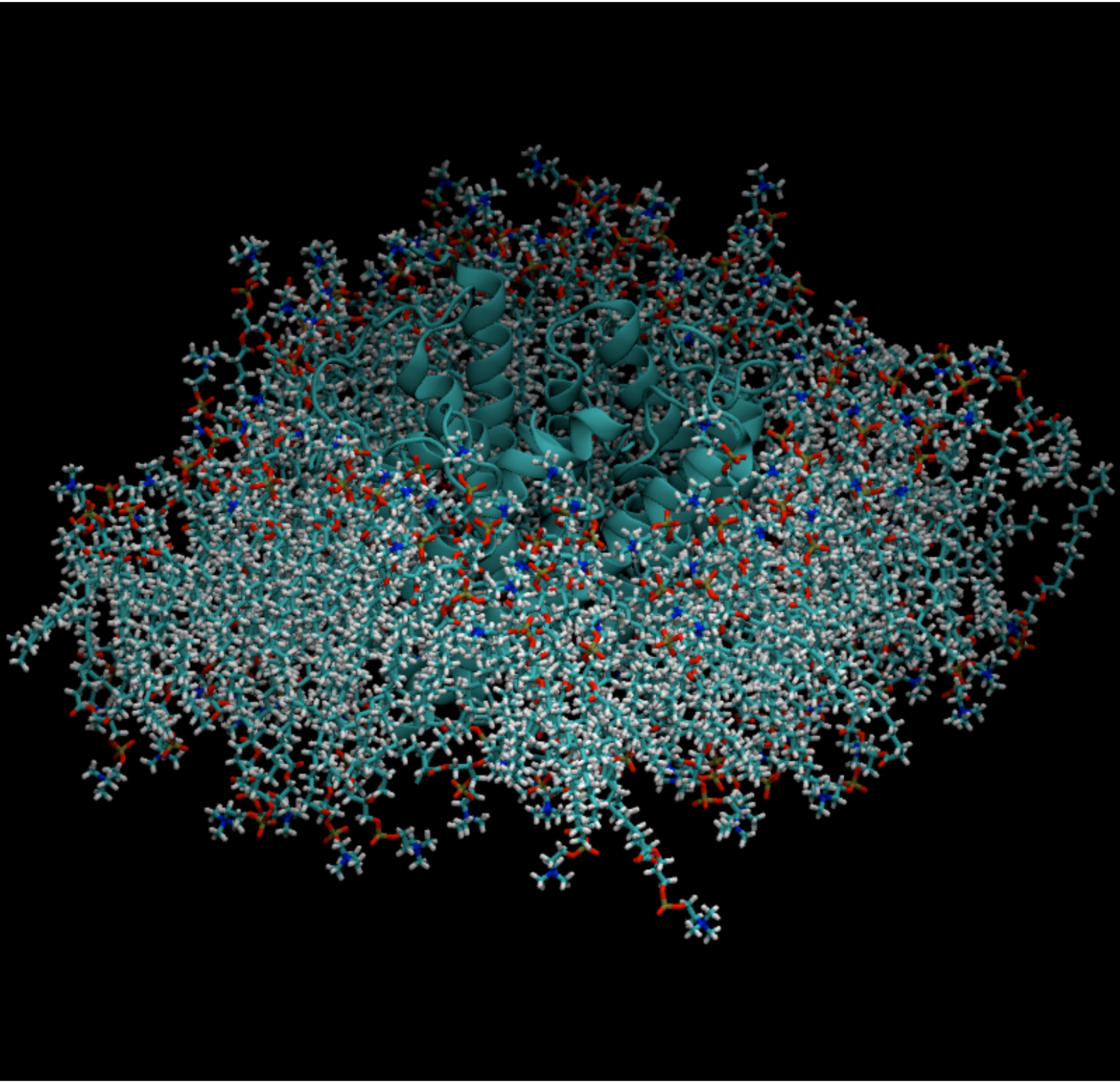
AmtB structure

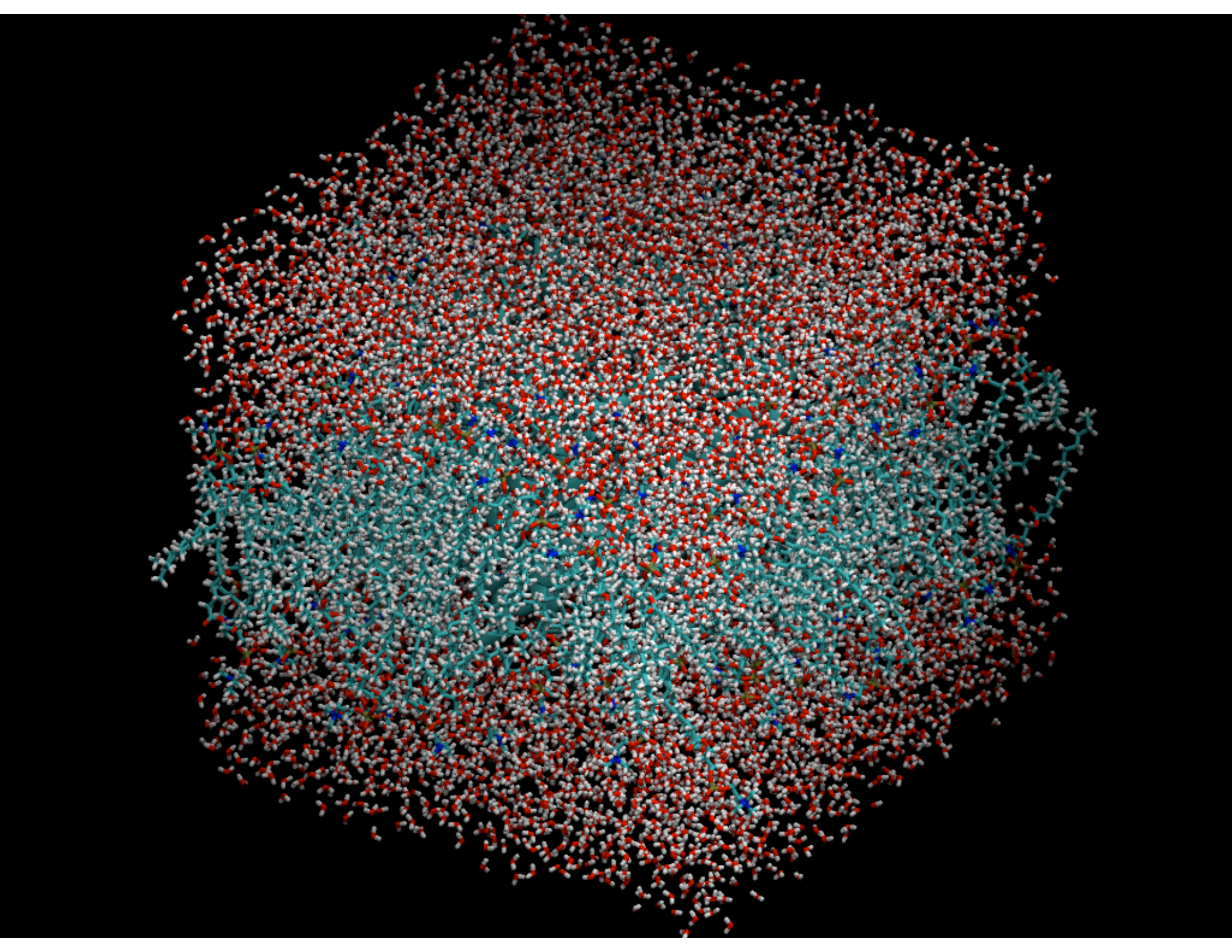


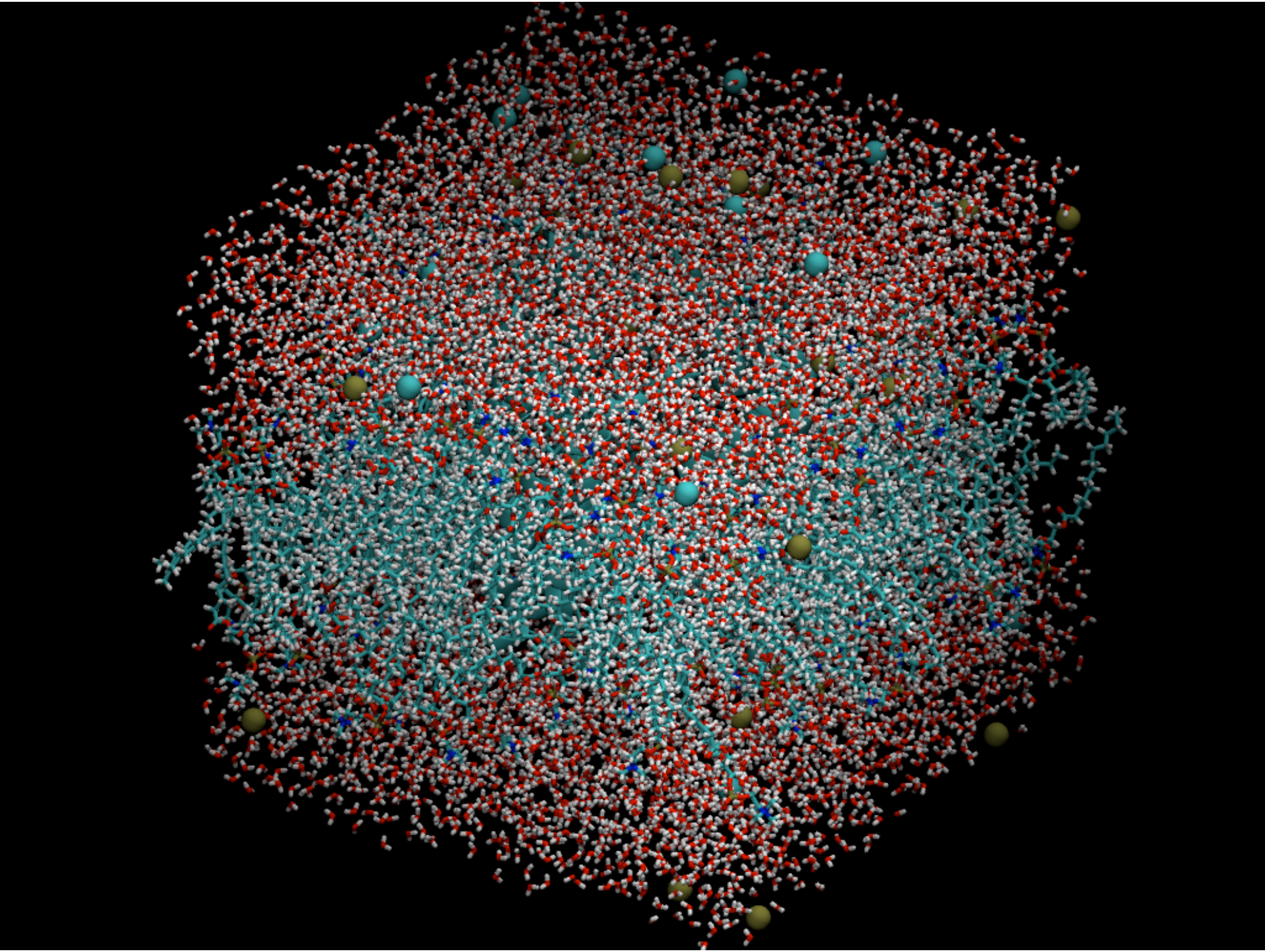
AmtB crystal structure

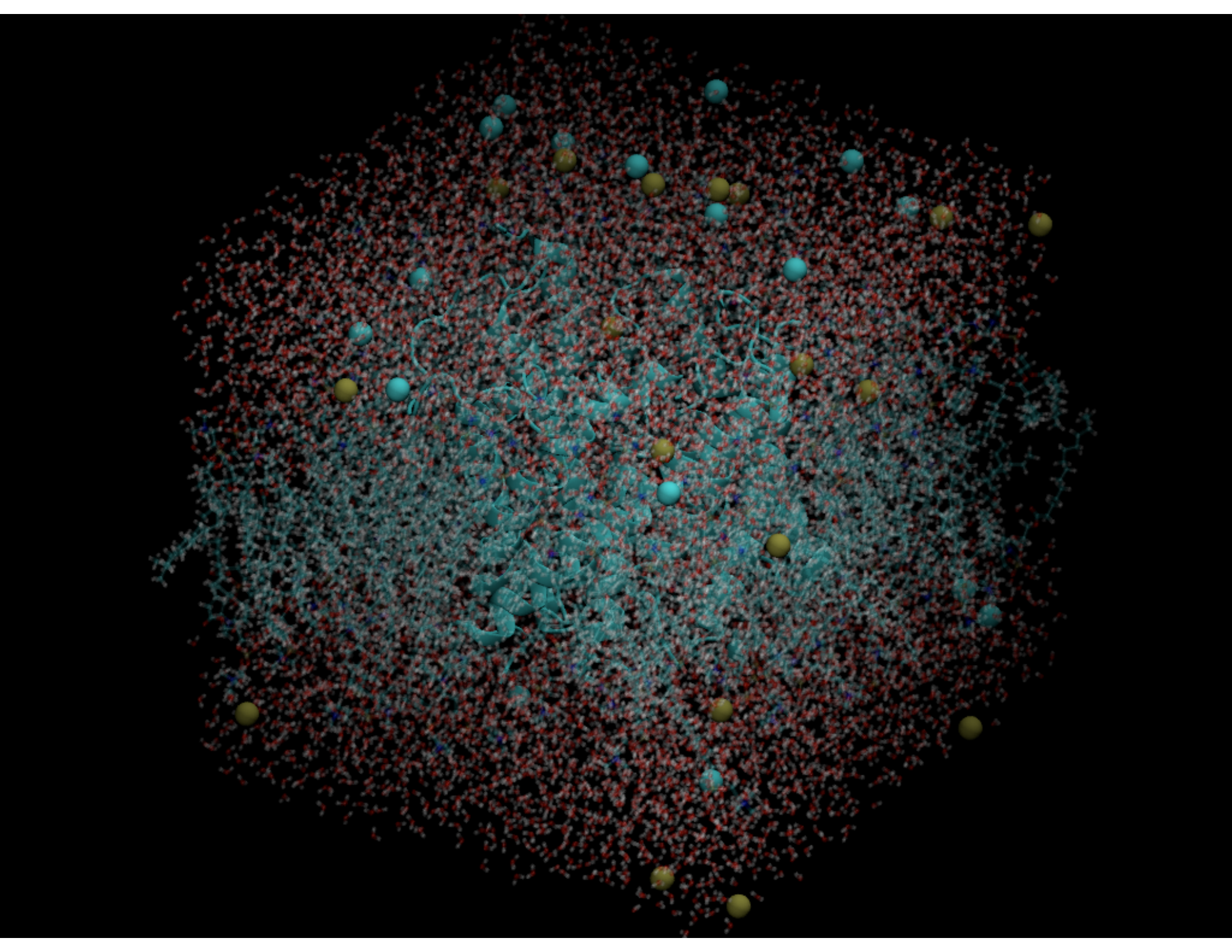
(Zheng et al., *PNAS* 2004, **101**, 17090)

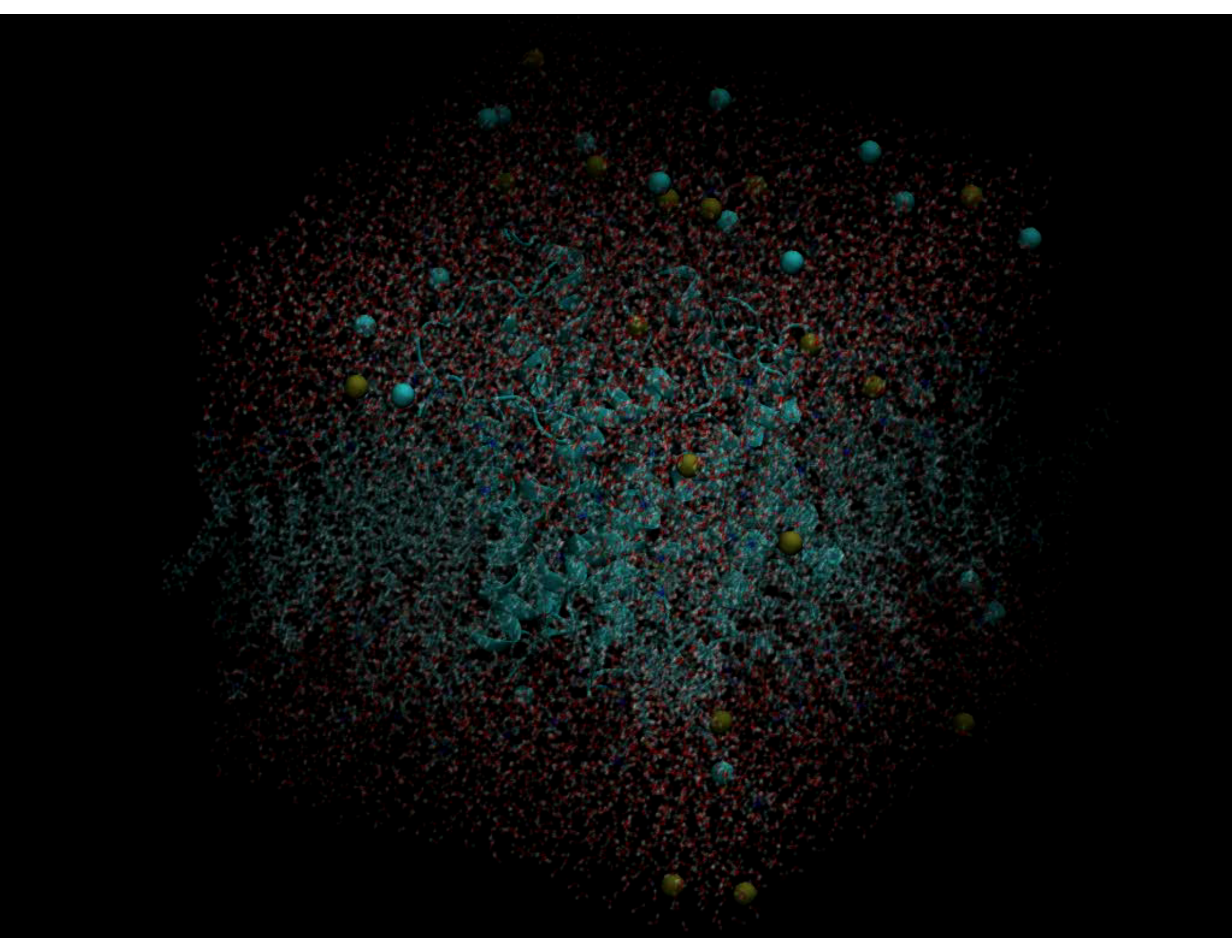












Debye–Waller factors (a.k.a. B-factors)

To compare with experimental B-factors, we have to account for crystal disorder, which is not present in the simulations.

$$B_i = \frac{8\pi}{3} \left\langle |\Delta \vec{r}_i|^2 \right\rangle + B_{\mathrm{W}}$$

$$\Delta \vec{r}_i = \vec{r}_i - \langle \vec{r}_i \rangle$$

Example: KcsA

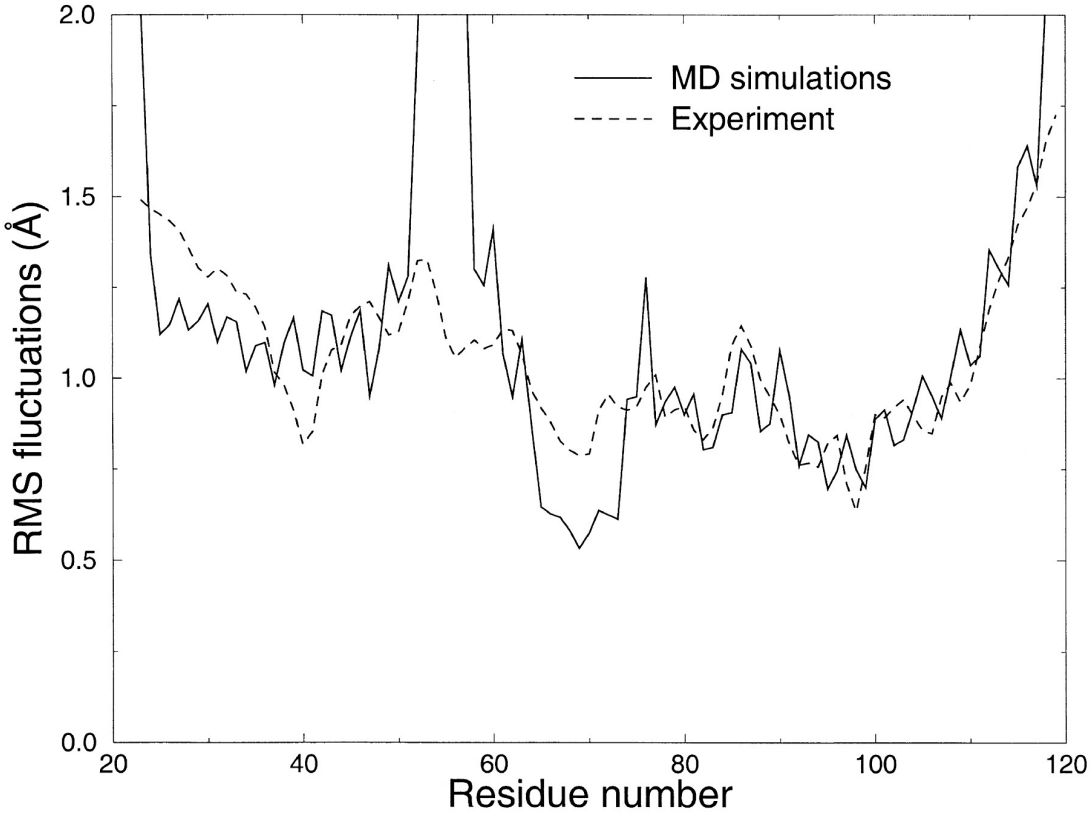


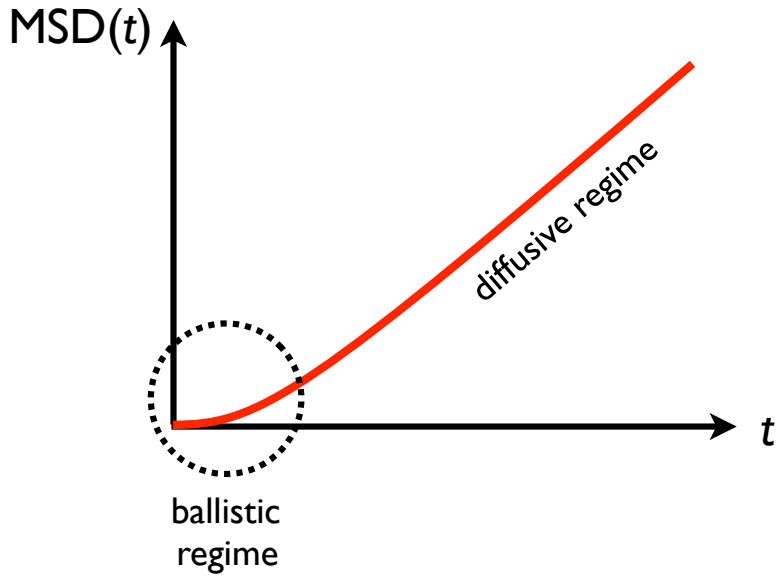
Figure from :
Bernèche and Roux, *Biophys. J.* **78**, 2900–2917 (2000)

Diffusion coefficients

(for a molecule in a homogeneous environment)

$$\mathrm{MSD}(t) = \left\langle |\mathbf{r}(t) - \mathbf{r}(0)|^2 \right\rangle$$

$$D = \lim_{t\to\infty} \frac{1}{6t} \mathrm{MSD}(t)$$



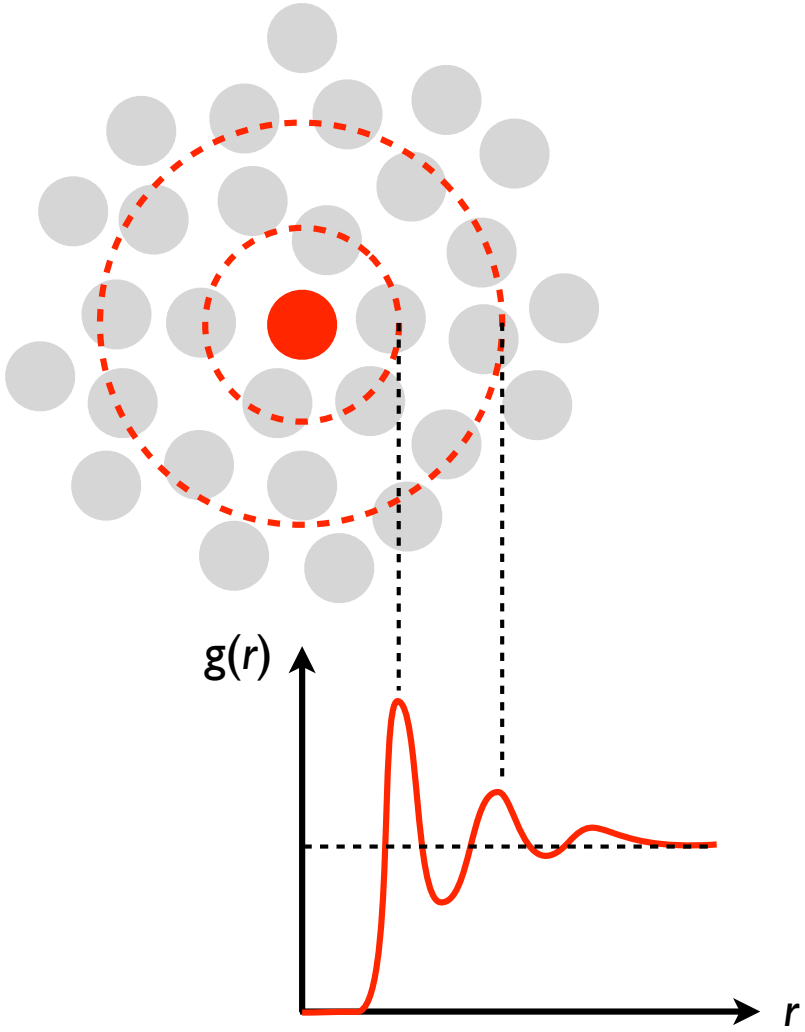
We usually calculate the MSD by assuming that the fluctuations are *stationary* :

$$\mathrm{MSD}(t) = \left\langle |\mathbf{r}(t+\tau) - \mathbf{r}(\tau)|^2 \right\rangle$$

$$= \frac{1}{N_\tau} \sum_{\tau}^{N_\tau} |\mathbf{r}(t+\tau) - \mathbf{r}(\tau)|^2$$

Radial distribution functions (RDFs)

(between pairs of atoms)



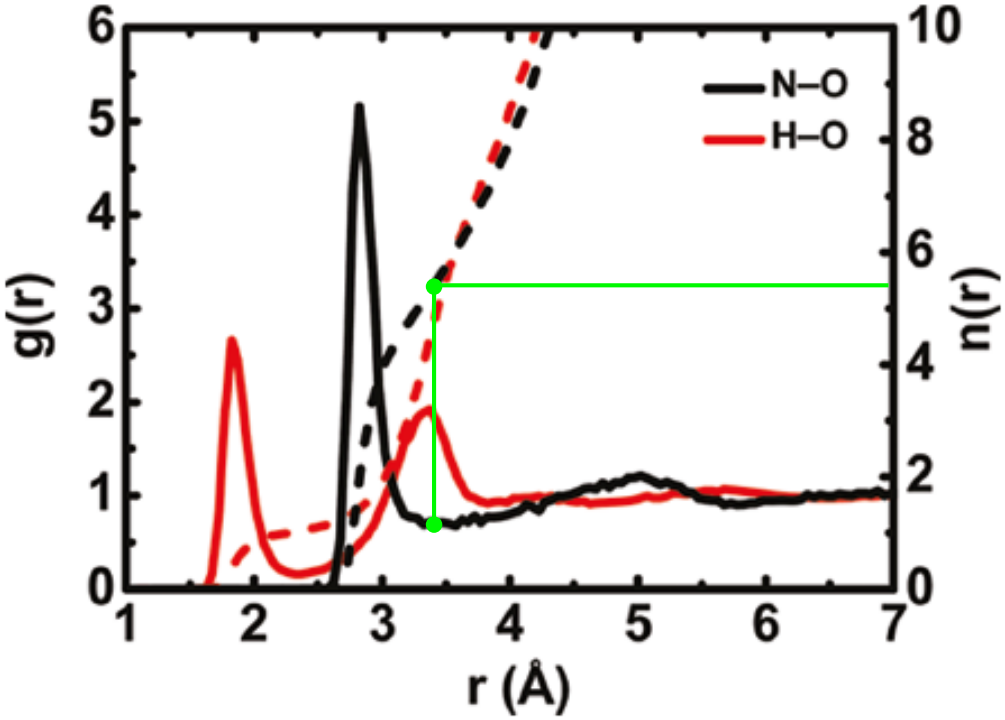


Figure 4. N–O and H–O radial distribution functions (solid lines, scale on left) and running integration numbers (dashed lines, scale on right) of NH_4^+ in water at 298.15 K.

Figure from :

Orabi and Lamoureux, *J. Chem. Theory Comput.* **8**, 182–193 (2012)