

Systematic coarse graining

Berk Hess

What is coarse graining?

For this talk I will define coarse graining as reducing the number of interactions for calculating the “potential”

Examples:

- Electronic density model $\longrightarrow$ classical force field
- Explicit solvent $\longrightarrow$ implicit solvent
- All-atom model $\longrightarrow$ united-atom model
- Atomistic model $\longrightarrow$ “chemical moiety model”

Why use coarse graining?

Example:

from 100000 atoms, 100 parameters

to 1000 particles, 10 parameters

Advantages:

- speeds up simulations
- easier to understand the physics
- possibly less over-fitting

Disadvantages:

- accurate force fields still required (and often not available)
- dynamics becomes inaccurate (faster, but how much?)

Reducing the number of degrees of freedom

Reducing the number of interactions usually goes together with reducing the number of degrees of freedom.

But one can reduce the number of degrees of freedom without reducing the number of interactions.

Example: torsional dynamics with an all atom model

Only reducing the number of degrees of freedom:

- Monte Carlo: smaller conformational space $\Rightarrow$ faster search
- Molecular Dynamics: more constraints $\Rightarrow$ slower

Why coarse grain?

Reasons for using a coarse grained model (vs. atomistic):

- For the system size I am interested in I can not reach the required time scale
- I am not interested in a specific chemical system, but in generic properties
- Simplified analysis of results

Speed-up of coarse graining:

- Cheaper potential calculation (factor 1-1000)
- Larger time step (factor 1-10)
- Smoother free energy landscape (artifact) (factor 1-1000)

Exact coarse graining

Given a potential V and a state point NpT :

$$\rho(\mathbf{R}) = \exp \left(-\frac{V(\mathbf{R})}{k_B T} \right)$$

We want to split the system in “interesting” coordinates $\mathbf{R}_a$ and “non-interesting” coordinates $\mathbf{R}_b$:

$$\rho_a(\mathbf{R}_a) = \int_{\mathbf{R}_b, (NpT)} \rho(\mathbf{R}) = \int_{\mathbf{R}_b, (NpT)} \exp \left(-\frac{V(\mathbf{R})}{k_B T} \right)$$

$\mathbf{R}_a$ can be any set of coordinates:

- coordinates of atoms
- centers of mass of molecules / groups of atoms
- distances between atoms
- ...

Exact coarse graining

We can convert ρ_a to a free energy W_a :

$$\begin{aligned} W_{a,NpT}(\mathbf{R}_a) &= -k_B T \log \rho_a(\mathbf{R}_a) \\ &= -k_B T \log \int_{\mathbf{R}_b, (NpT)} \exp \left(-\frac{V(\mathbf{R})}{k_B T} \right) \\ &= V_a(\mathbf{R}_a) - k_B T \log \int_{\mathbf{R}_b, (NpT)} \exp \left(-\frac{V(\mathbf{R}) - V_a(\mathbf{R}_a)}{k_B T} \right) \end{aligned}$$

Range of validity

Coarse grained “potential” (really free energy) W :

$$W_{a,NpT}(\mathbf{R}_a) = -k_B T \log \int_{\mathbf{R}_b, (NpT)} \exp \left(-\frac{V(\mathbf{R})}{k_B T} \right)$$

Strictly speaking $W_{a,NpT}$ can **only** be used at $N = N_0$, $p = p_0$, $T = T_0$

Composition: e.g. $N_0 = \{N_a = 100, N_b = 1000\}$:

$W_{a,NpT}$ only valid for $N_a/N_b = 1/10$

Pressure / volume

Because $W_{a,NpT}$ is not a potential

virial “pressure” $\neq$ thermodynamic pressure

simulate at volume $V_0 = \langle V \rangle_{NpT}$

Dynamics

Removing degrees of freedom also removes their friction and noise contributions.

Leaving out friction and noise will speed up the dynamics

Friction and noise can have complex time and space correlations, which are related through the dissipation-fluctuation theorem.

Some friction and noise can be put back in using for instance stochastic dynamics (Langevin dynamics)

Stochastic dynamics (in Gromacs) has no time or space correlation.

n-body expansion

Without loss of generality we can expand W in n -body terms:

$$\begin{aligned} W_{a,NpT}(\mathbf{R}_a) &= \sum_{i < j} W_2(r_{ij}) \\ &+ \sum_{i < j < k} W_3(r_{ij}, r_{ik}, r_{jk}) \\ &+ \sum_{i < j < k < l} W_4(r_{ij}, r_{ik}, r_{il}, r_{jk}, r_{jl}, r_{kl}) \\ &+ \dots \end{aligned}$$

summing over all a -particles

Pair potential approximations

Infinite dilution approximation:

$$\begin{aligned}
 W_{a,NpT}(\mathbf{R}_a) &\approx \sum_{i < j} W_2(r_{ij}) \\
 &+ \sum_{i < j < k} W_3(r_{ij}, r_{ik}, r_{jk}) \\
 &+ \sum_{i < j < k < l} W_4(r_{ij}, r_{ik}, r_{il}, r_{jk}, r_{jl}, r_{kl}) \\
 &+ \dots
 \end{aligned}$$

Effective pair potential:

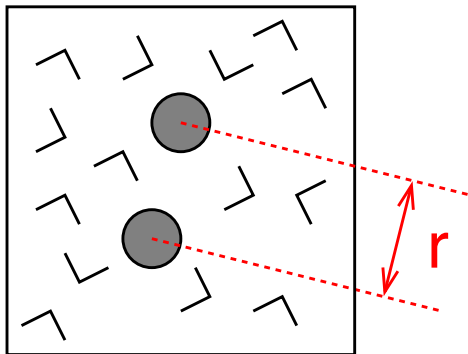
$$W_{a,NpT}(\mathbf{R}_a) = \left. \begin{aligned} &\sum_{i < j} W_2(r_{ij}) \\ &+ \sum_{i < j < k} W_3(r_{ij}, r_{ik}, r_{jk}) \\ &+ \sum_{i < j < k < l} W_4(r_{ij}, r_{ik}, r_{il}, r_{jk}, r_{jl}, r_{kl}) \\ &\dots \end{aligned} \right\} \approx \sum_{i < j} W_p(r_{ij})$$

Obtaining infinite dilution potentials

The infinite dilution “potential” is a potential of mean force (PMF).

A PMF is the free energy along one degree of freedom, averaging over all other degrees of freedom.

The mean force between two molecules can be determined by constraining the distance between the two molecules in an MD simulation and monitoring the constraint force.



$$\text{PMF}(r) = \int_{\infty}^r \langle F_{\text{constraint}} \rangle_s ds$$

PMF: an entropy correction

The PMF as described contains the kinetic entropy term of two masses rotating at fixed distance.

For a constraint in a d -dimensional system this term is:

$$G(r) = -k_B T \log(r^{d-1}) = -(d-1) k_B T \log(r)$$

The entropic force pushing the two molecules apart is:

$$F(r) = -\frac{dG}{dr} = (d-1) k_B T \frac{1}{r}$$

Defining effective pair potential

- Use an analytical form
example: Martini
example: in Mainz: WCA potential (LJ cut at minimum)
- Require all pair correlations to match the reference:
unique solution for the effective pair potentials

Matching all pair correlations:

advantage: all pairs correlations correct

disadvantage: what happens to higher order correlations?

Effec. pair pot.: iterative Boltzmann inversion

- Simulate the detailed system
- Determine the RDF $g_{ref}(r)$ between the coarse-grained centers

First guess for the effective potential W :

$$W^0(r) = -k_B T \log(g_{ref}(r))$$

Determine the RDF: $g^0(r)$

Iteration:

$$W^i(r) = W^{i-1}(r) + k_B T \log \left(\frac{g^{i-1}(r)}{g_{ref}(r)} \right)$$

Determine the RDF: $g^i(r)$

Converged when $g^i(r)$ is “close enough” to $g_{ref}(r)$

Effective pair potential: inverse Monte Carlo

Sometimes inverse Boltzmann has convergence problems, especially for multi component systems

Example: three component system, components a and b and solvent

three target RDF's: $g_{aa}(r)$, $g_{bb}(r)$, $g_{ab}(r)$

three effective pair potentials: $W_{aa}(r)$, $W_{bb}(r)$, $W_{ab}(r)$

All six are interlinked

Effective pair potential: inverse Monte Carlo

Sometimes inverse Boltzmann has convergence problems, especially for multi component systems

Example: three component system, components a and b and solvent
three target RDF's: $g_{aa}(r)$, $g_{bb}(r)$, $g_{ab}(r)$

three effective pair potentials: $W_{aa}(r)$, $W_{bb}(r)$, $W_{ab}(r)$

All six are interlinked

A stable, efficient method: inverse Monte Carlo

- divide all $g_{ref}(r)$ and $W(r)$ in spatial bins
- determine all correlations between all bins in g_{ref}
- do Monte Carlo (or MD) with the current guess for the potentials
- adjust the $W_{??}(r)$ bin values taking into account the correlations

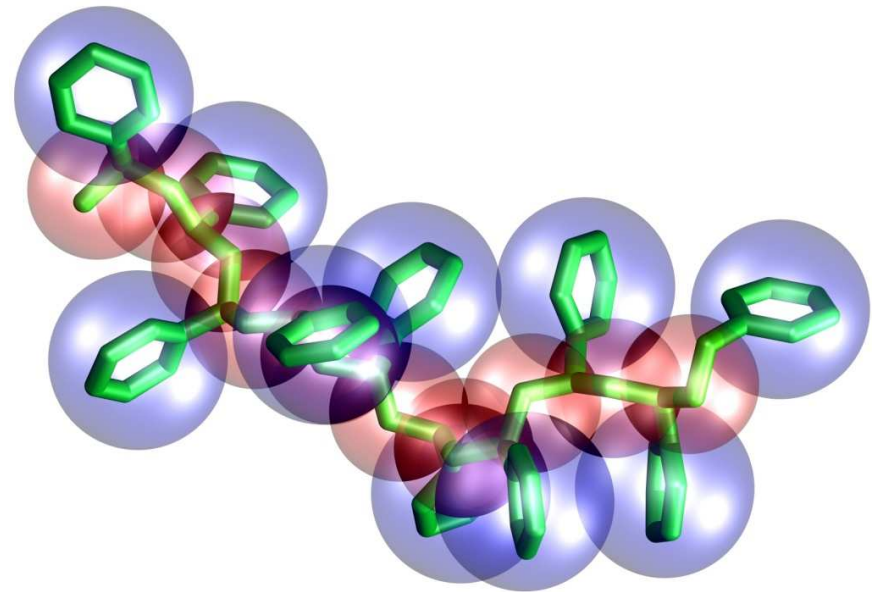
Systematic coarse-graining of polymers

Non-bonded interactions:
iterative Boltzmann inversion
for a liquid

Bonded interactions:
direct Boltzmann inversion

Possible issues:

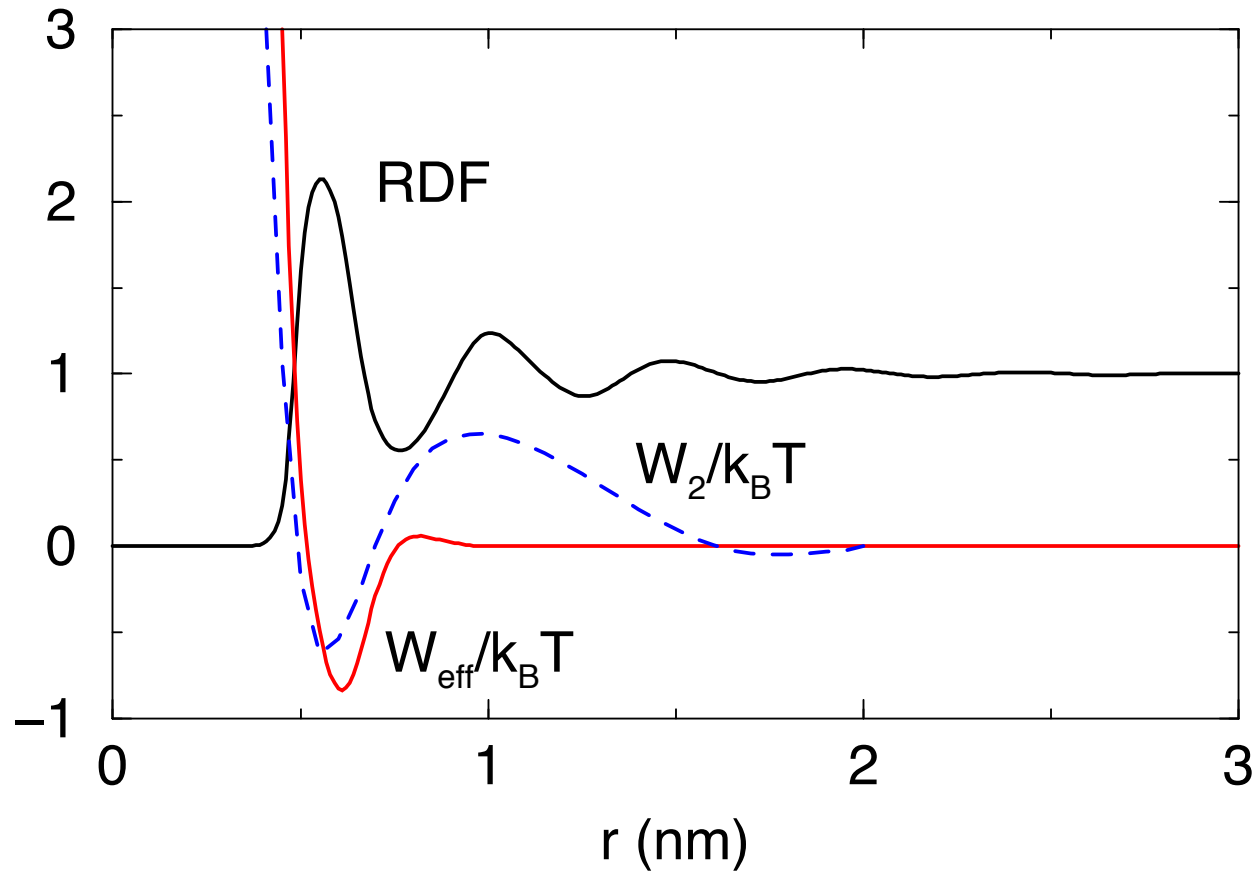
- (de)coupling of bonded interactions
- what is bonded and what is non-bonded along the chain?



polystyrene

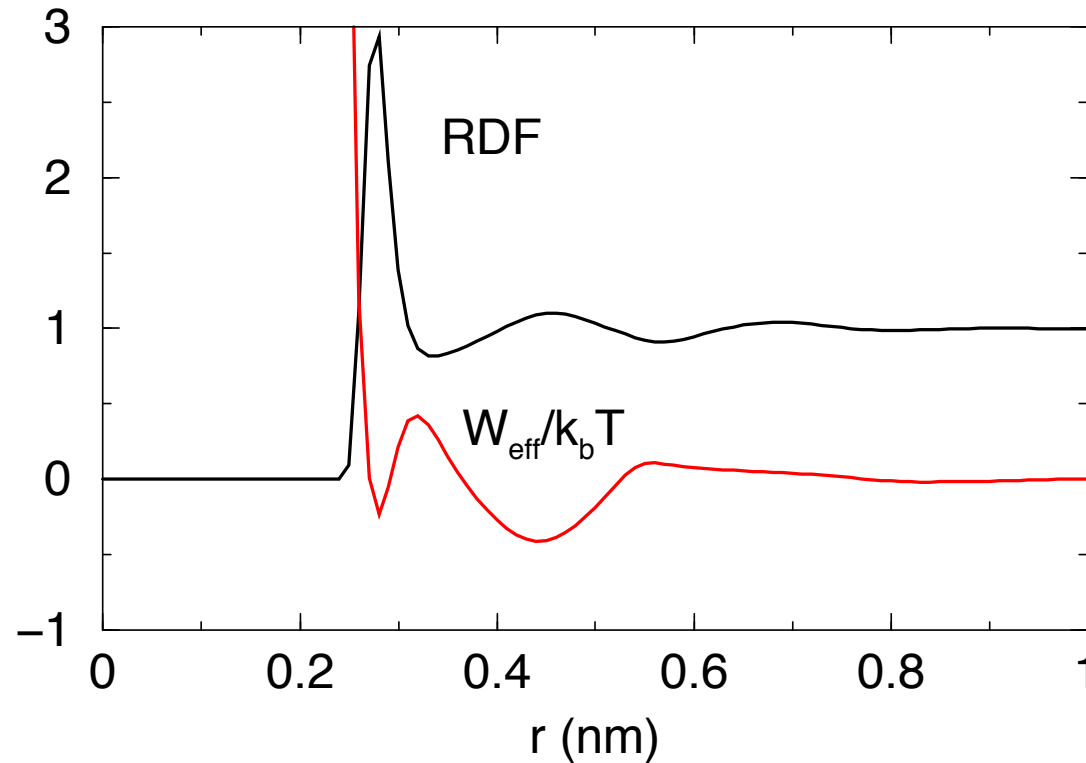
Effective pair potential: liquid benzene

one bead: one benzene molecule



Effective pair potential: SPC/E water

one bead: one water molecule



Strong multi-body effects squeezed into a pair potential:

- Bad higher order structure
- extreme virial pressures: 9000 bar

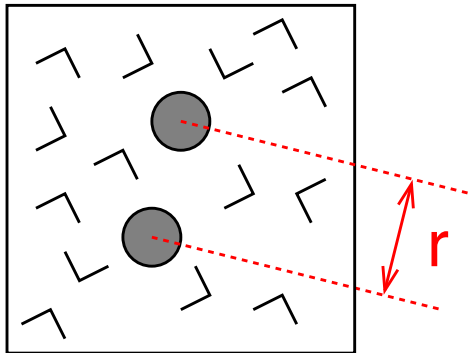
Example: ions in implicit solvent

Ions in explicit water $\longrightarrow$ ions in implicit water

Capturing the multi-body effects in a single parameter

Calculating PMF's between ions

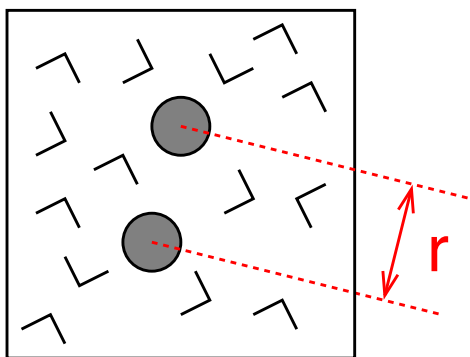
The mean force between two ions can be determined by constraining the distance between the two ions in an MD simulation and monitoring the constraint force.



$$\text{PMF}(r) = \int_{\infty}^r \langle F_{\text{constraint}} \rangle_s ds$$

Calculating PMF's between ions

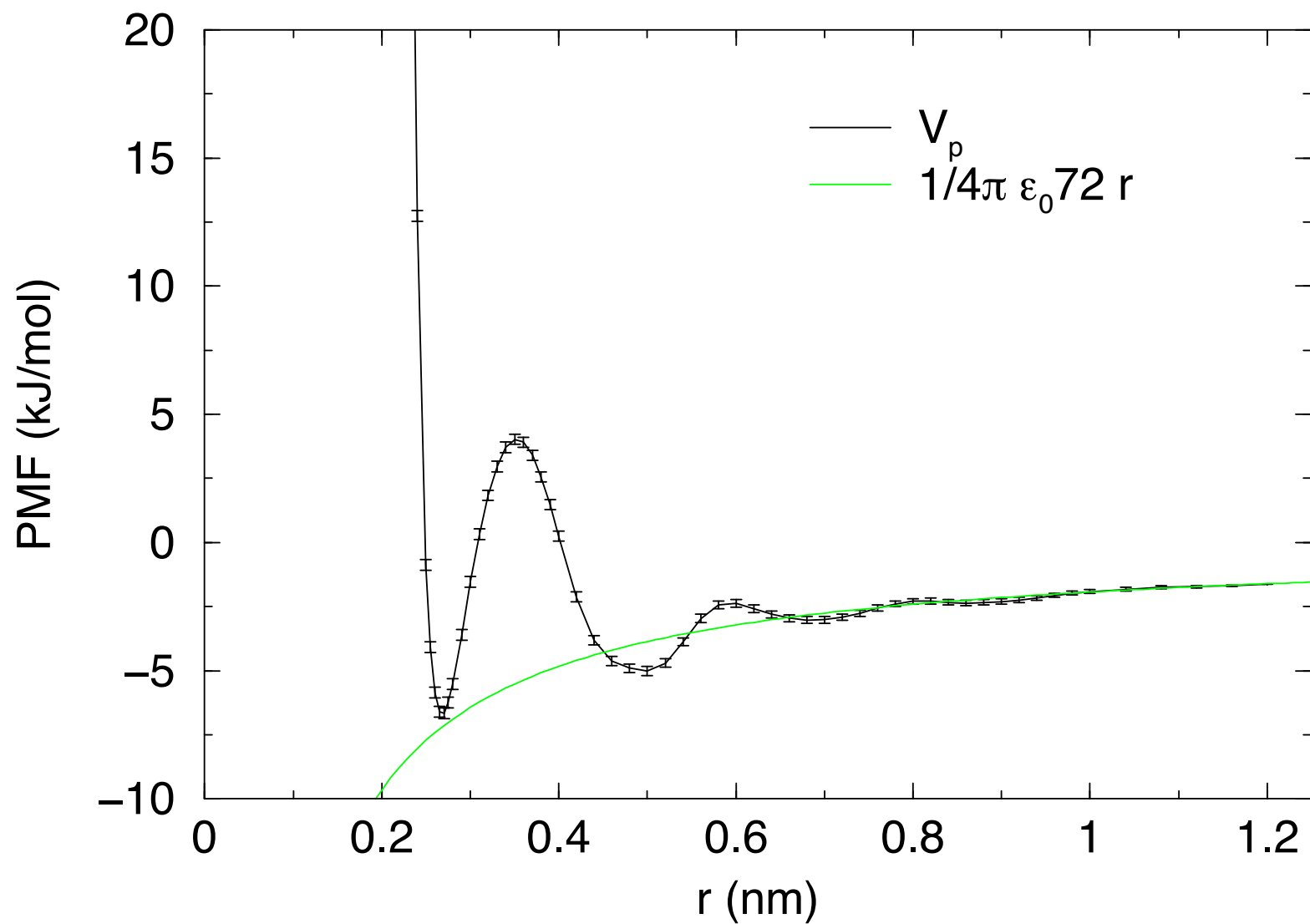
The mean force between two ions can be determined by constraining the distance between the two ions in an MD simulation and monitoring the constraint force.



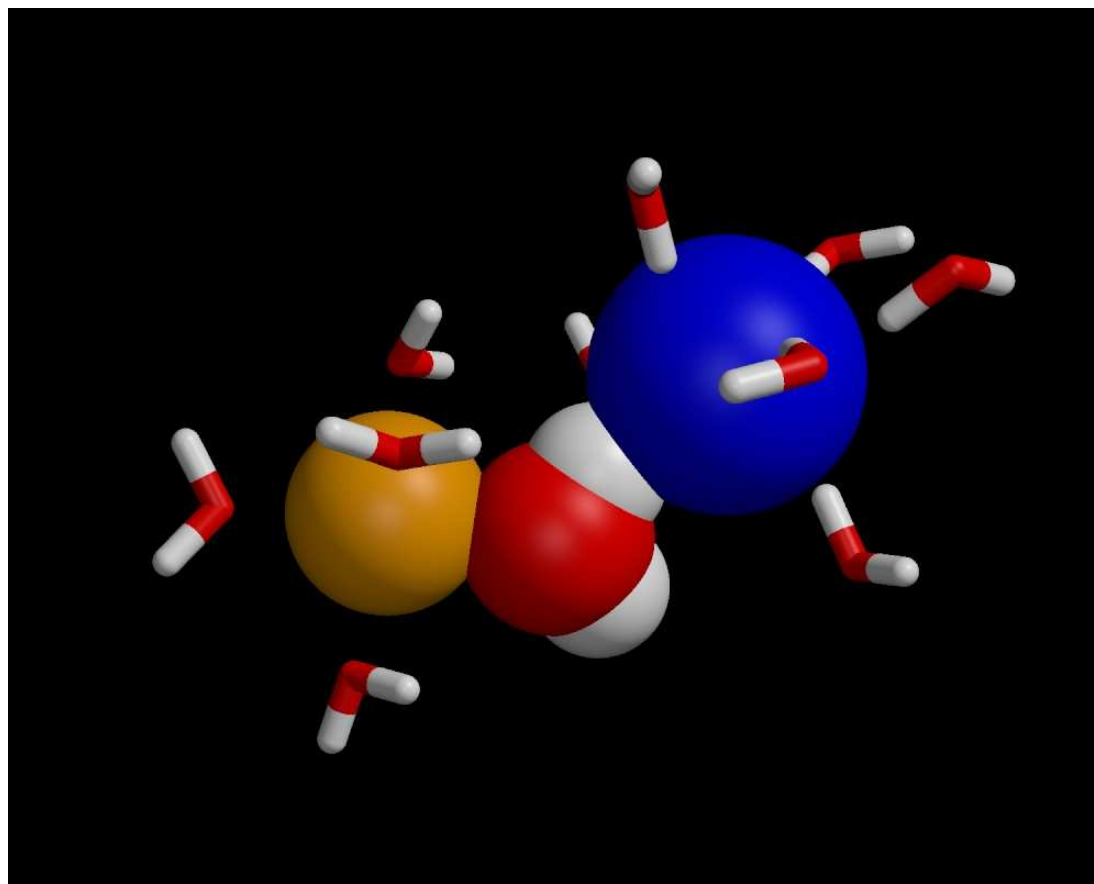
$$\text{PMF}(r) = \int_{\infty}^r \langle F_{\text{constraint}} \rangle_s ds$$

- 1 ion pair, 1000 SPC/E water molecules (SPC/E: $\epsilon = 72$)
- periodic unit cell, NPT ensemble, PME electrostatics
- GROMACS simulation package

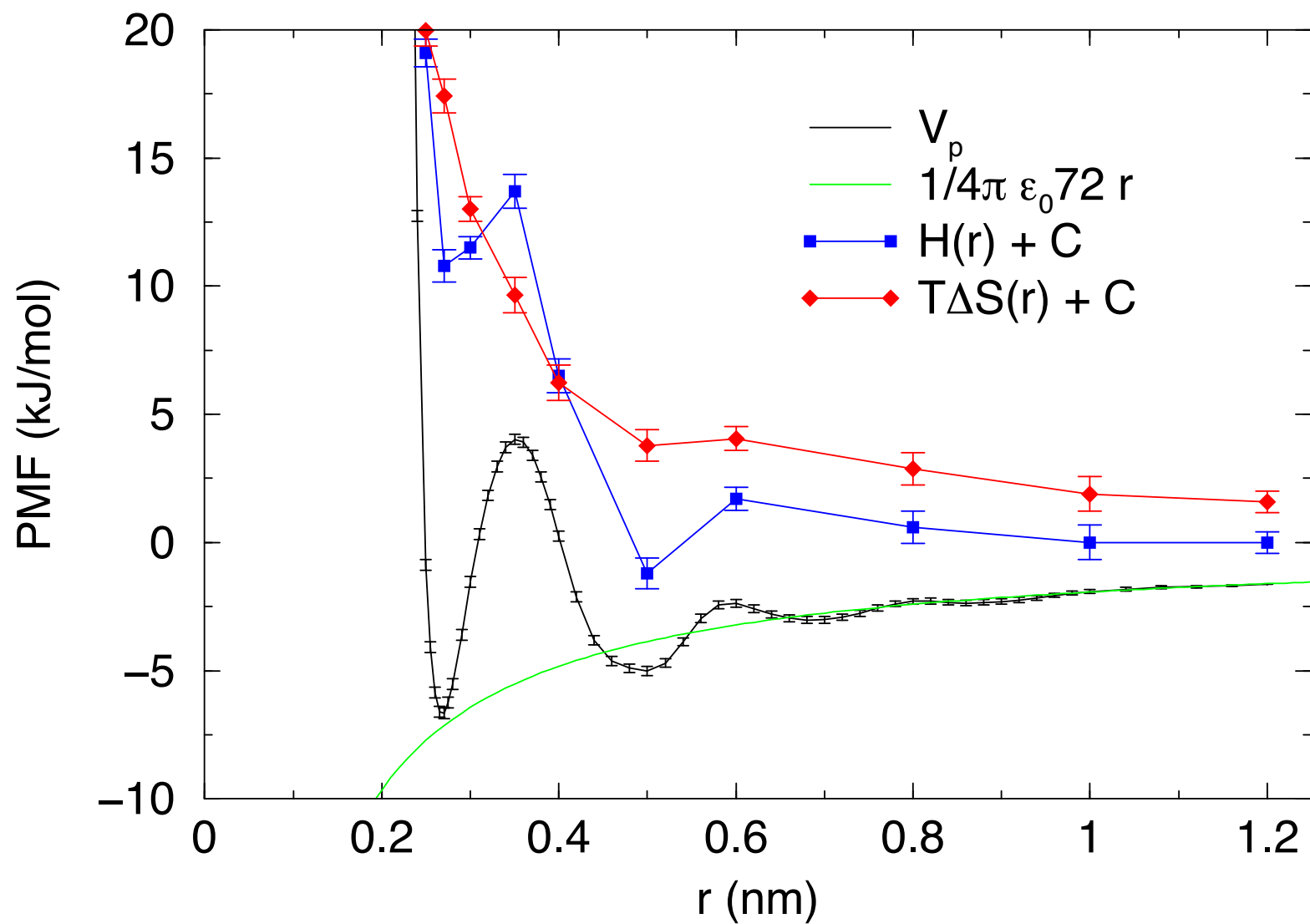
$\text{Na}^+ - \text{Cl}^-$ solvent-averaged interaction



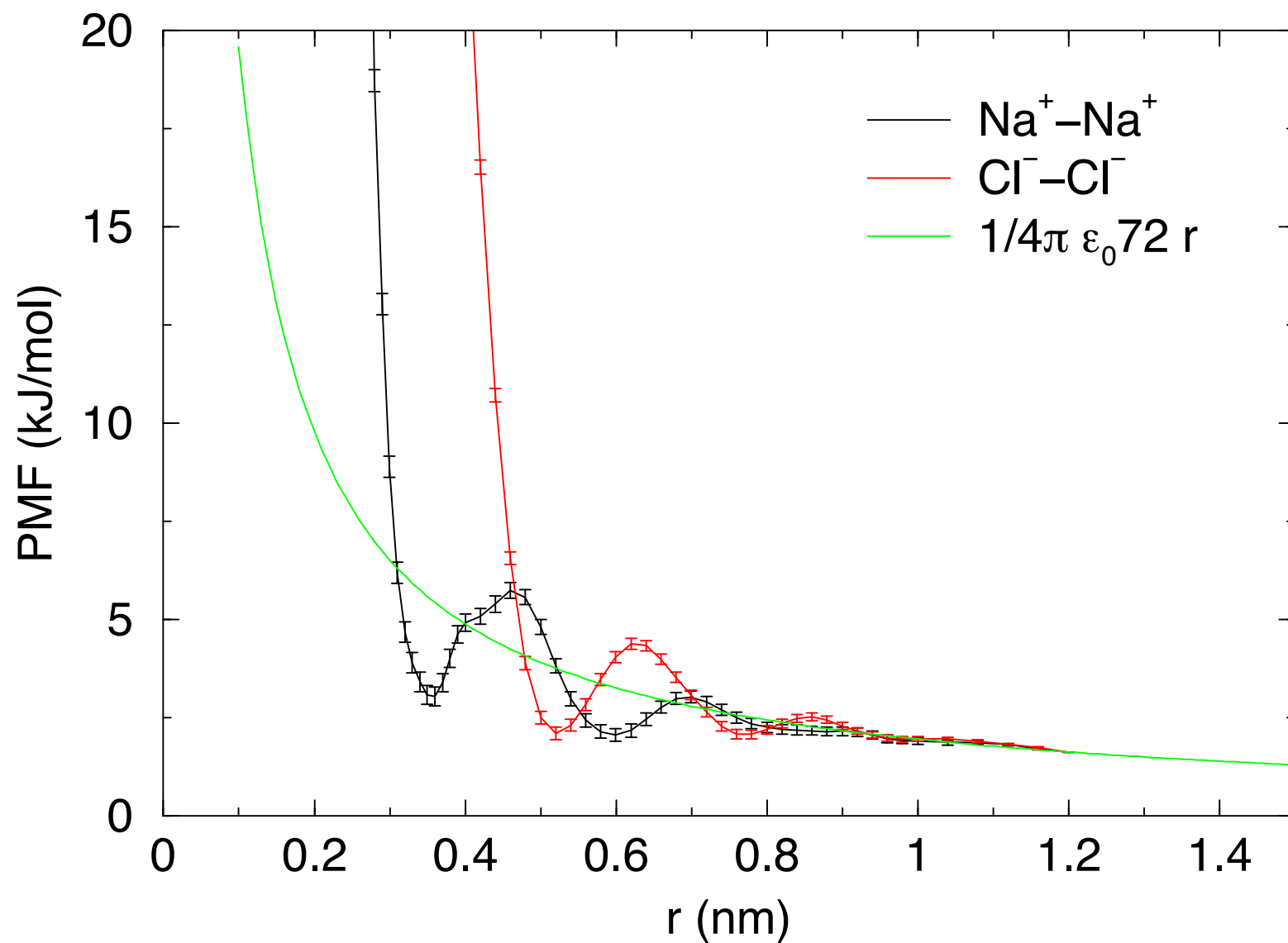
$\text{Na}^+ - \text{Cl}^-$ solvent separated pair



$\text{Na}^+ - \text{Cl}^-$ solvent-averaged interaction



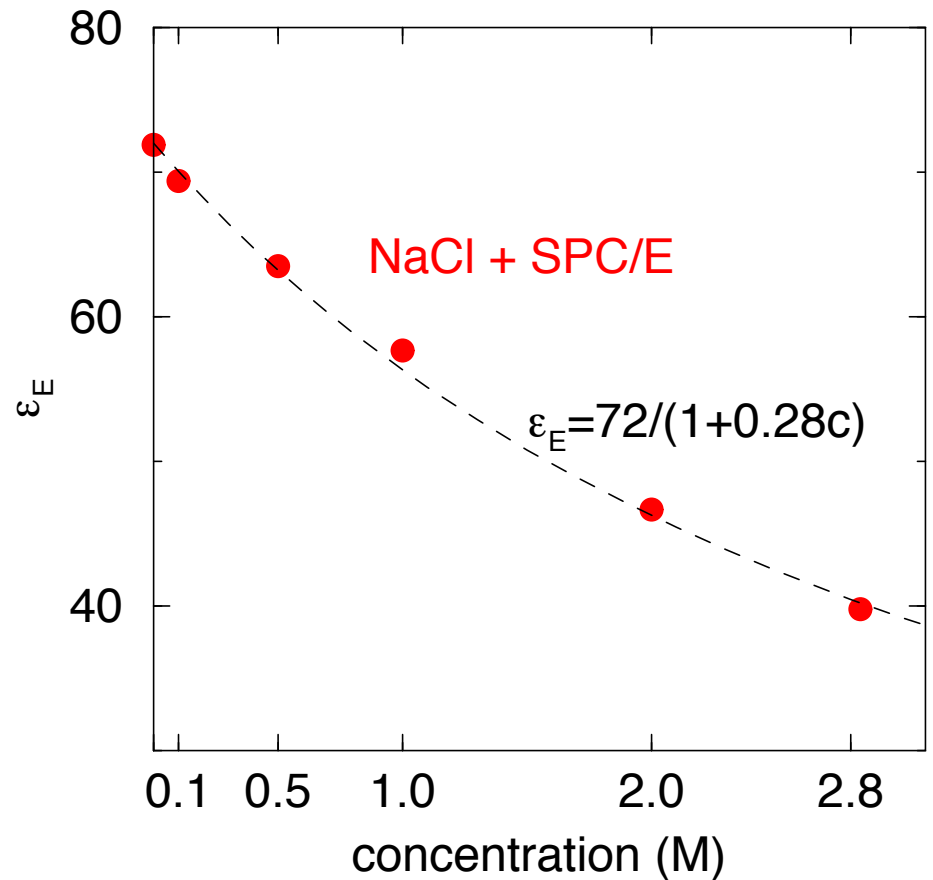
Eq. charged solvent-averaged interaction



Solute dependent dielectric permittivity

The solute or concentration dependent dielectric permittivity ϵ_E is given by the dipole fluctuation of the solvent only

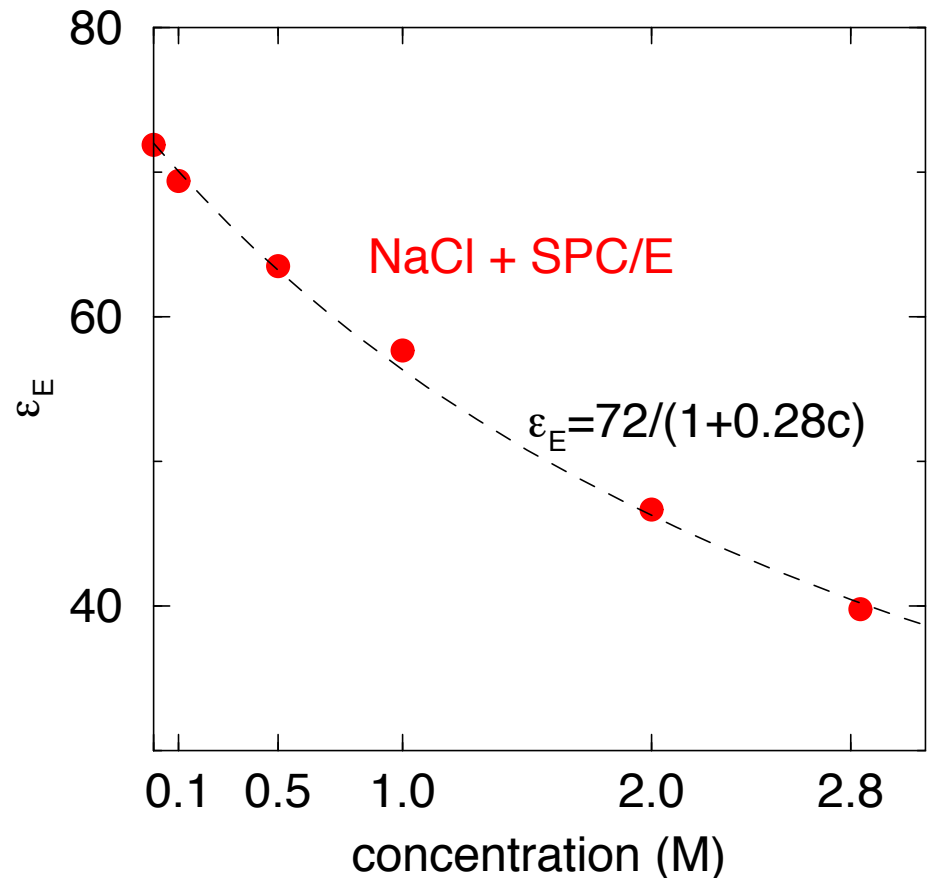
Levesque et al. JCP 72, 1887 (1980)



Solute dependent dielectric permittivity

The solute or concentration dependent dielectric permittivity ϵ_E is given by the dipole fluctuation of the solvent only

Levesque et al. JCP 72, 1887 (1980)

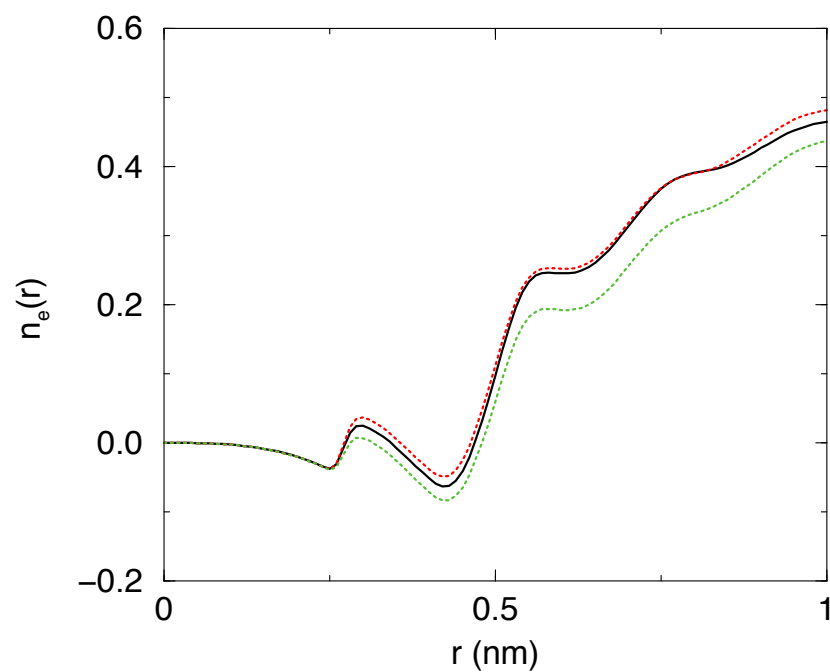
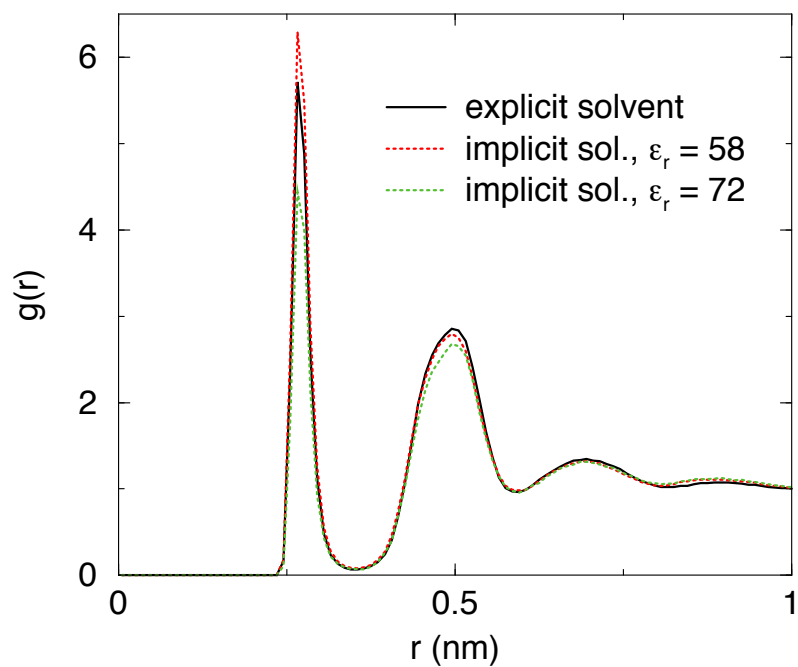


$$V_c(r, c) = PMF(r) + \frac{q_1 q_2}{4 \pi \epsilon_0} \left(\frac{1}{\epsilon_E(c)} - \frac{1}{\epsilon_E(0)} \right) \frac{1}{r}$$

Hess et al. PRL 96, 147801 (2006)

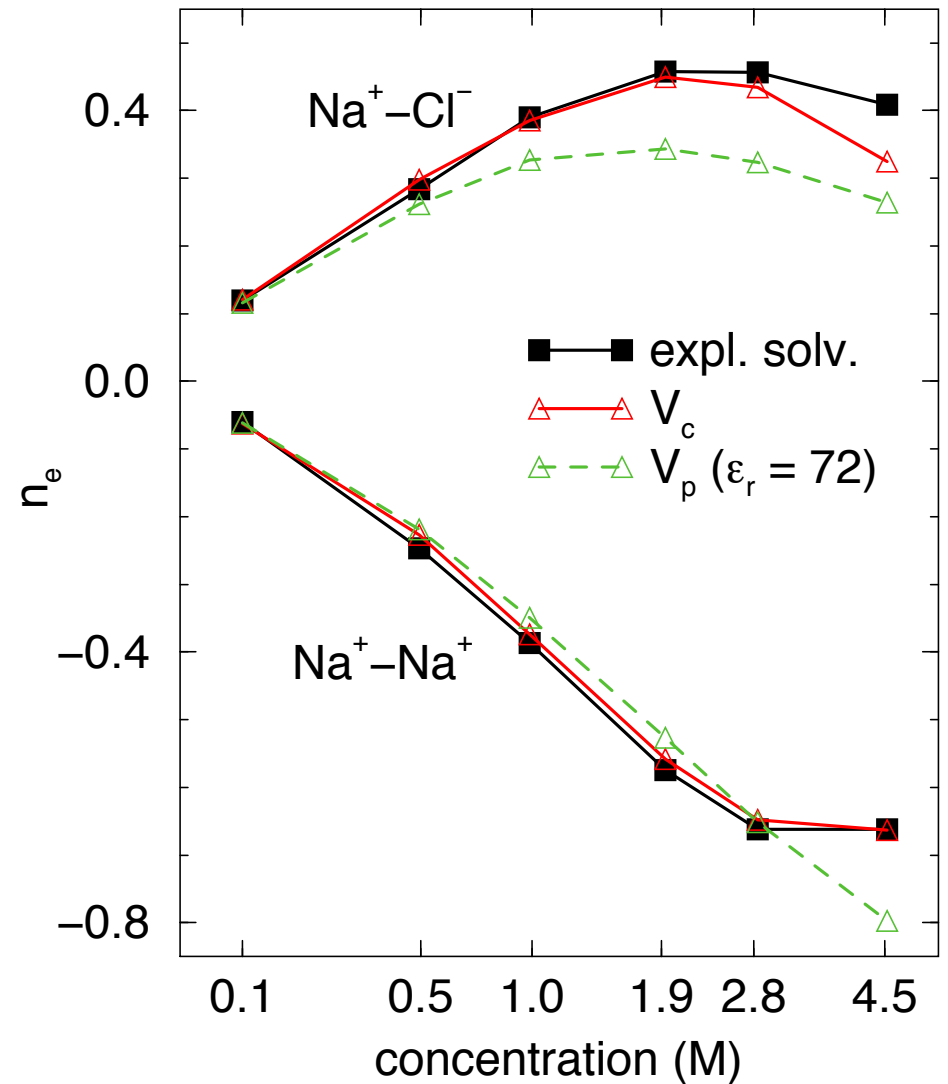
Implicit solvent results

$$n_e(r) = \rho 4 \pi \int_0^r [g(s) - 1] s^2 ds$$



Excess coordination (up to 0.8 nm)

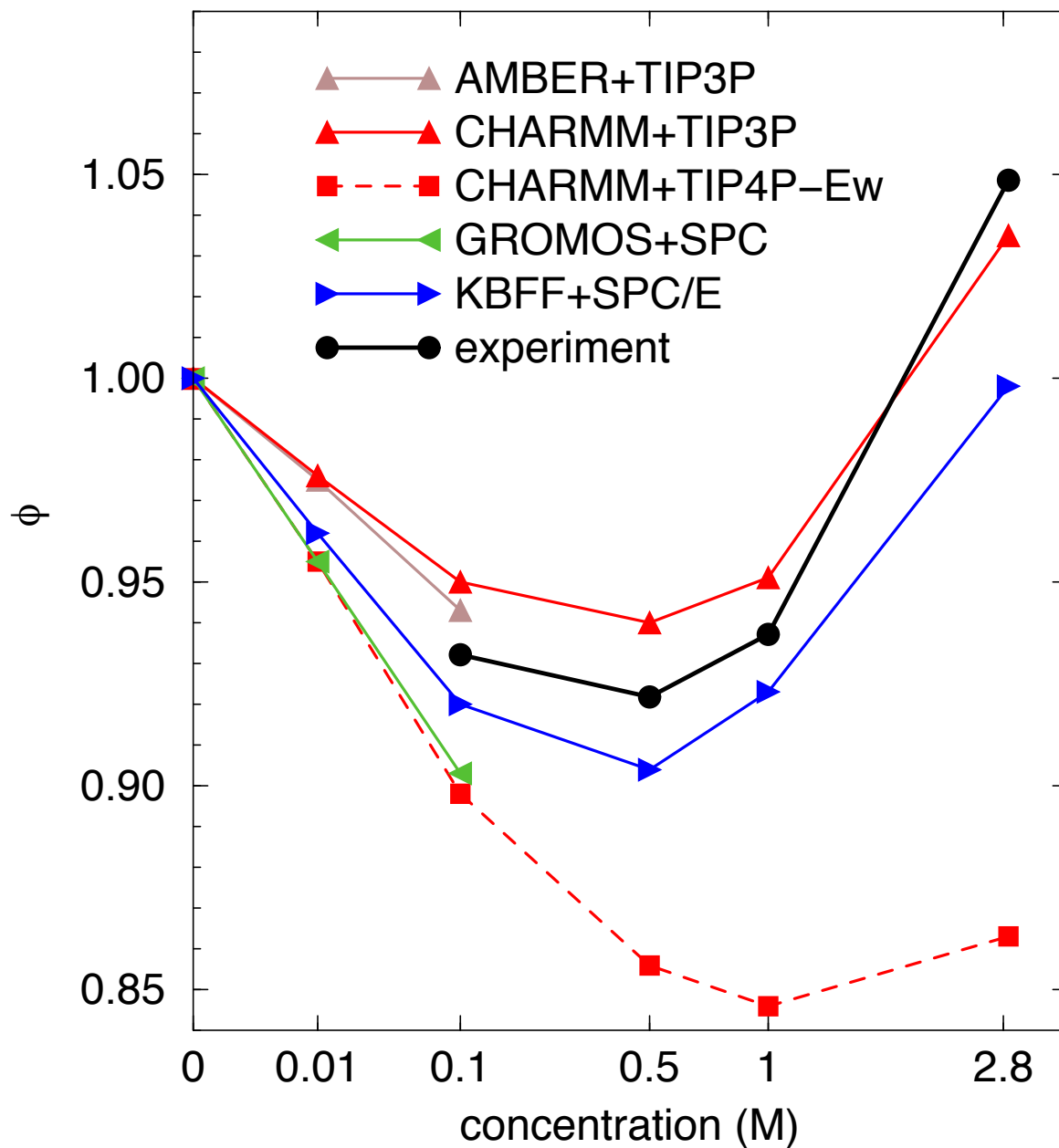
$$n_e(r) = \rho 4 \pi \int_0^r [g(s) - 1] s^2 ds$$



Direct comparison with experiment

When using
implicit solvent:

$$\phi = \frac{P}{P_{ideal}}$$



Gromacs user tabulated functions

You can use any functional shape for bonded and non-bonded interaction through user tables.

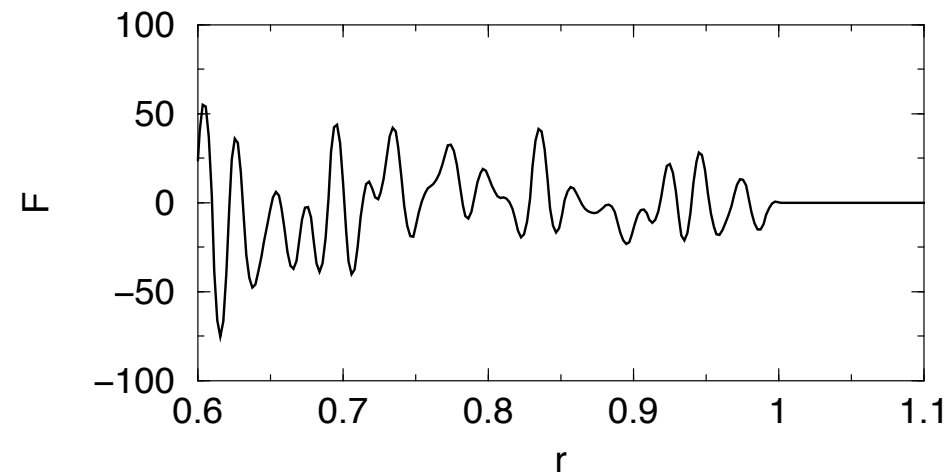
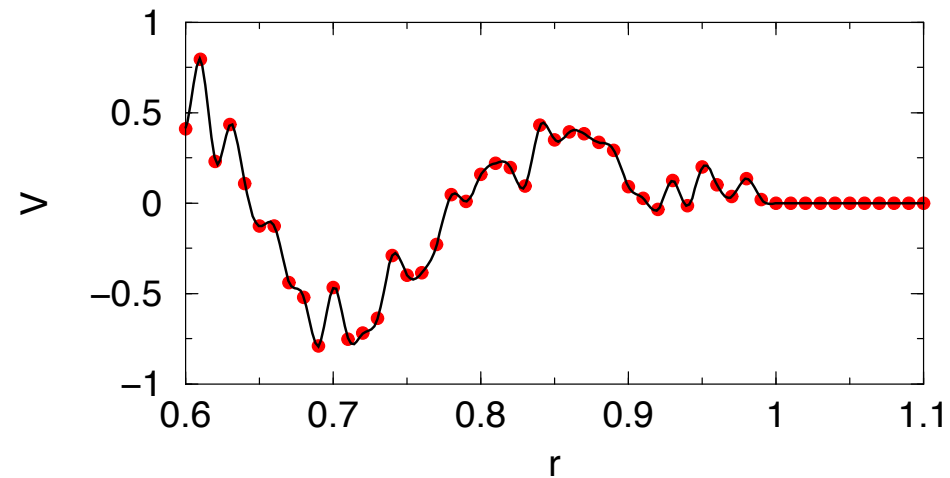
The tables in Gromacs use cubic spline interpolation, this gives continuous and consistent potential and forces

For splines you need to specify V and $F = -dV/dr$ at each table point

Table files are simply ascii files which are read by mdrun

Derivative generation

Hidden feature of mdrun:
generate potential derivatives using Akima splines



Exercise

Derive an implicit solvent model for methane in water

Determine a PMF, make a tabulated potential function

Run implicit solvent simulations and compare with explicit solvent