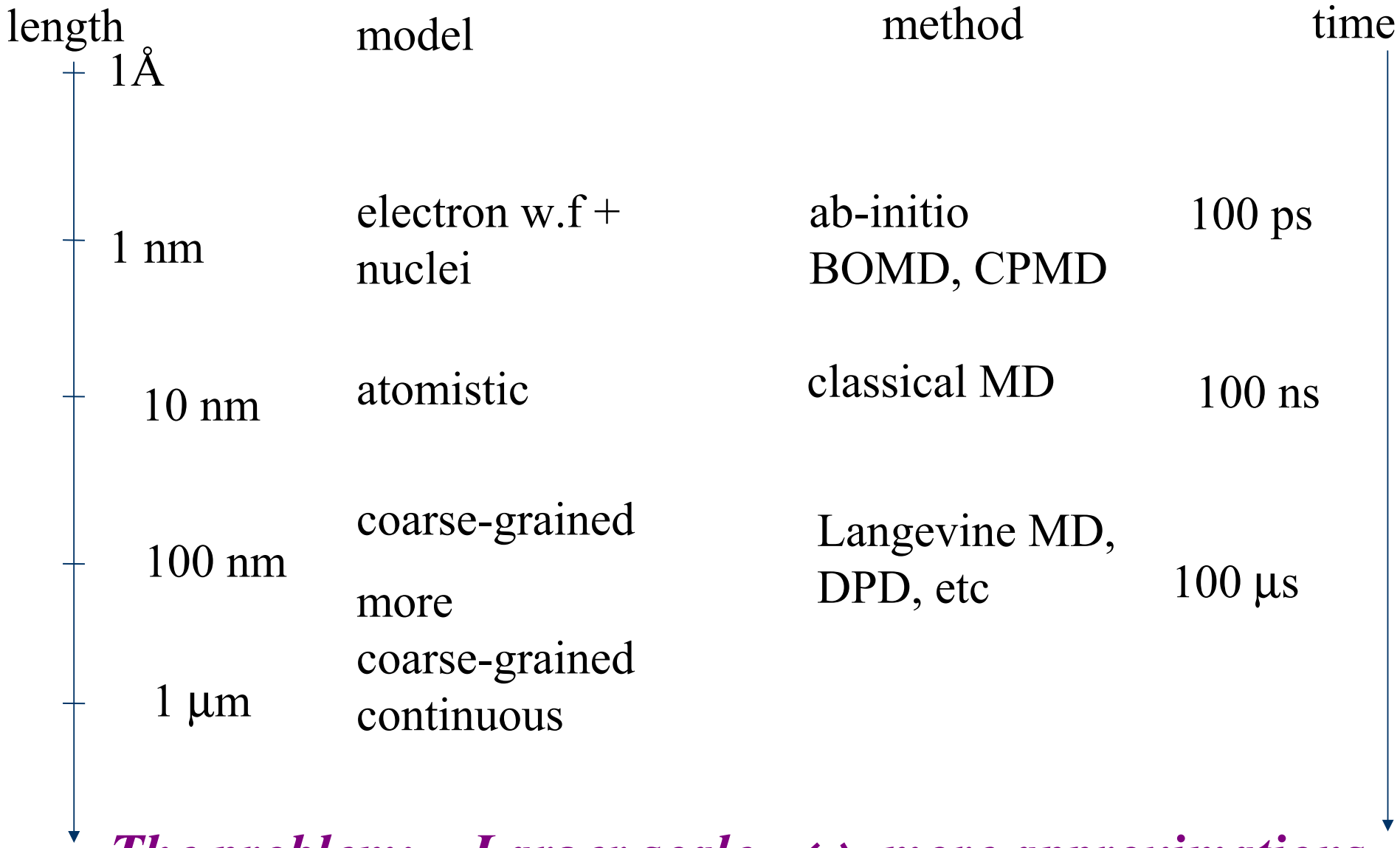
Systematic Coarse-Graining of Molecular Models by the Inverse Monte Carlo: Theory, Practice and Software

Alexander Lyubartsev

Division of Physical Chemistry
Department of Material and Environmental Chemistry
Stockholm University

Modeling Soft Matter:
Linking Multiple Length and Time Scales
Kavli Institute of Theoretical Physics, UCSB, Santa Barbara 4-8 June 2012

Soft Matter Simulations



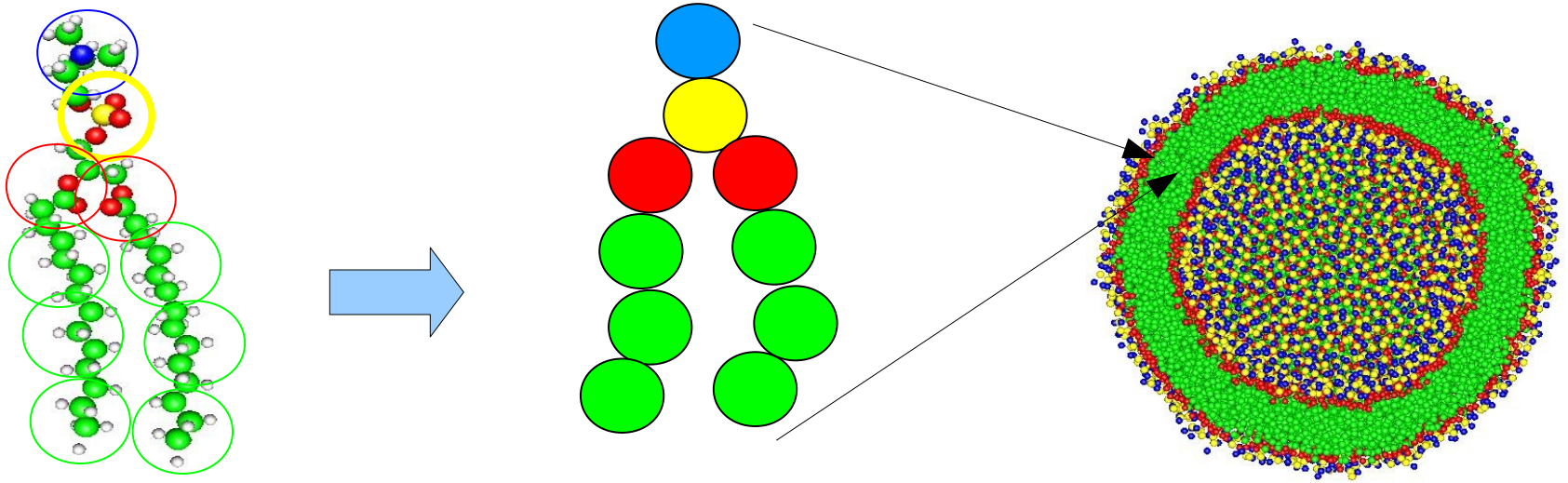
The problem: Larger scale ⇔ more approximations

Coarse-graining – an example

Original size - 900K

Compressed to 24K

Coares-graining: reduction degrees of freedom



All-atom model
118 atoms

Coarse-grained model
10 sites

Large-scale
simulations

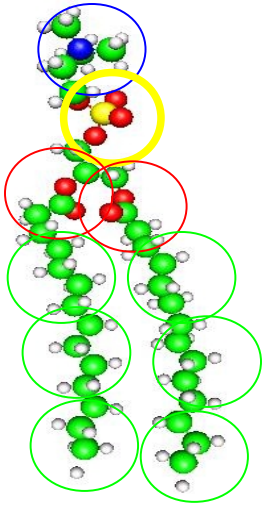
We need to:

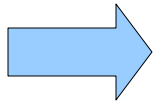
1) Design Coarse-Grained mapping: specify the important degrees of freedom
2) For “important” degrees of freedom we need interaction potential

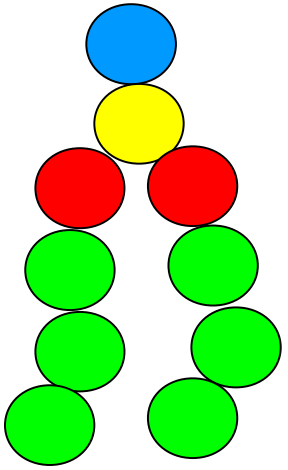
Question: what is the interaction potential for the coarse-grained model?

Formal solution:
N-body mean force potential

Original (FG = fine grained system)







$H^{FG}(r_{1,}r_{2,}\ldots,r_n)$

$n = 118$

$R_j = \theta(r_{1,}\ldots r_n)$

$j = 1,\ldots,10$

Usually, centers of mass of selected molecular fragments

Partition function :

$$Z = \int \prod_{i=1}^{n} dr_i \exp\left(-\beta H_{FG}(r_{1,}\ldots, r_n)\right) =$$

$$= \int \prod_{i_1}^{n} dr_i \prod_{j=1}^{N} dR_j \,\delta\left(R_j - \theta_j(r_{1,}\ldots r_n)\right)\exp\left(-\beta H_{FG}(r_{1,}\ldots, r_n)\right) =$$

$$= \int \prod_{j=1}^{N} dR_j \exp\left(-\beta H_{CG}(R_{1,}\ldots, R_N)\right)$$

where $\beta = \dfrac{1}{k_B T}$

$$H_{CG}(R_{1,}\ldots, R_N) = -\frac{1}{\beta}\ln \int \prod_{i=1}^{n} dr_i \prod_{j=1}^{N} \delta\left(R_j - \theta_j(r_{1,}\ldots, r_n)\right)\exp\left(-\beta H_{FG}(r_{1,}\ldots r_n)\right)$$

is the effective N-body coarse-grained potential = potential of mean force = free energy of the non-important degrees of freedom

N-body potential of mean force (CG Hamiltonian) $H_{CG}(R_1,...R_N)$:

Structure: all structural properties are the same

for any $A(R_1,..R_N)$: $\langle A \rangle_{FG} = \langle A \rangle_{CG}$

Thermodynamics: in principle yes (same partition function)

but: $H_{CG} = H_{CG}(R_1,...R_N, \beta, V)$

already N-body mean force potential is state point dependent
T-V dependence should be in principle taken into account in computing thermodynamic properties

Dynamics: can be approximated within Mori-Zvanzig formalism

(e.g within DPD: Eriksson et all, PRE, 77, 016707 (2008))

Problem with $H_{CG}(R_1,...R_N)$: simulation with N-body potential is unrealistic. ***We need something usable – e.g. pair potentials!***
(or any other preferably one-dimensional functions)

$$H_{CG}(R_1, R_2, \ldots, R_N) \approx \sum_{i>j} V_{ij}(R_{ij}) \qquad R_{ij} = |R_i - R_j|$$

How to approximate ?

- minimize the difference (Boltzmann averaged) : Energy matching
- minimize the difference of gradient (force) Force matching
- minimize the relative entropy
- ***provide the same canonical averages, e.g RDF-s***
  - Inverse MC
  - Iterative Boltzmann inversion

.. and may be some other properties "Newton inversion"

These approaches are in fact interconnected ....

The method

(A.P.Lyubartsev, A.Mirzoev, L.J. Chen, A.Laaksonen, Faraday Discussions, 144, 2010)

Assume: $H(\lambda_1,\lambda_2,...\lambda_M)$: Hamiltonian (potential energy) depending on a set of parameters $\lambda_1,\lambda_2,...\lambda_M$.

We wish to reproduce M canonical averages $\langle S_1\rangle,\langle S_2\rangle,...,\langle S_M\rangle$

Set of λ_α , $\alpha=1,\ldots,M \leftrightarrow$ Space of Hamiltonians

$\{\lambda_\alpha\}$ $\underset{\text{inverse}}{\overset{\text{direct}}{\rightleftharpoons}}$ $\{\langle S_\alpha\rangle\}$

In the vicinity of an arbitrary point in the space of Hamiltonians one can write:

$$\Delta \langle S_\alpha \rangle = \sum_\gamma \frac{\partial \langle S_\alpha \rangle}{\partial \lambda_\gamma} \Delta\lambda_\gamma + O(\Delta\lambda^2)$$

where

$$\frac{\partial \langle S_\alpha \rangle}{\partial \lambda_\gamma} = \left\langle \frac{\partial S_\alpha}{\partial \lambda_\gamma} \right\rangle - \beta \left(\left\langle \frac{\partial H}{\partial \lambda_\alpha} S_\gamma \right\rangle - \left\langle \frac{\partial H}{\partial \lambda_\alpha} \right\rangle \langle S_\gamma \rangle \right)$$

This allows us to solve the inverse problem iteratively

Algorithm:

Choose trial values $\lambda_\alpha^{(0)}$

↓

Direct MC or MD

↓

Calculate $\langle S_\alpha \rangle^{(n)}$ and differences

$\Delta\langle S_\alpha \rangle^{(n)} = \langle S_\alpha \rangle^{(n)} - S_\alpha^*$

↓

Solve linear equations system

↓

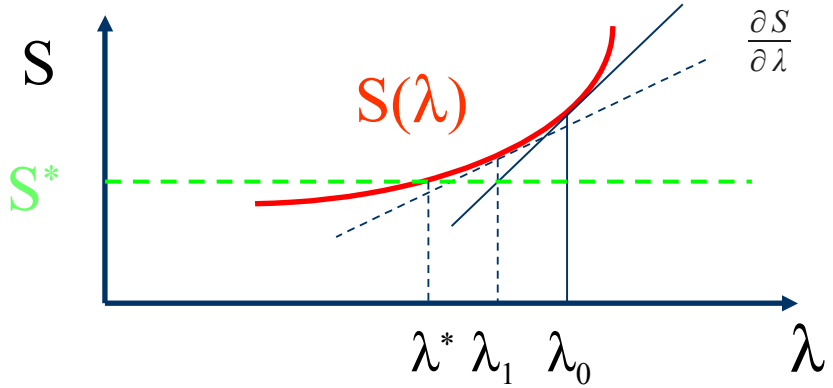
Obtain $\Delta\lambda_\alpha^{(n)}$

↓

New Hamiltonian: $\lambda_\alpha^{(n+1)} = \lambda_\alpha^{(n)} + \Delta\lambda_\alpha^{(n)}$

Repeat until convergence

Newton method of solving non-linear equation:



if no convergence, a regularization procedure can be applied:

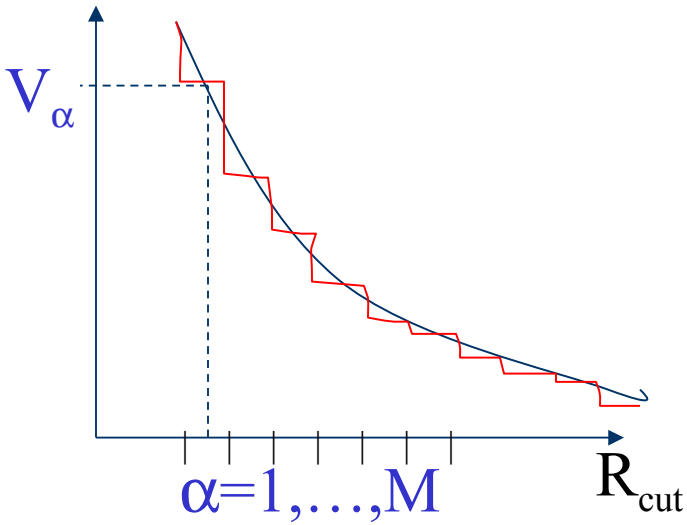
$\Delta\langle S_\alpha \rangle^{(n)} = a(\langle S_\alpha \rangle^{(n)} - S_\alpha^*)$

with $a<1$

Inverse Monte Carlo:

Reconstruction of pair potentials from RDFs

A.Lyubartsev, A.Laaksonen,Phys,Rev.E, 52, 1995)



We consider Hamiltonians in the form:

$$H = \sum_{\alpha} V_\alpha S_\alpha(r_i)$$

Any pair potential $H = \sum_{i,j} V(r_{ij})$ can be written in this way within the grid approximation::

$V_\alpha = V(R_{cut}\alpha/M)$ - potential within α-interval = λ- parameters

S_α - number of particle's pairs with distance between them within α-interval =

estimator of RDF: $$g(r_\alpha) = \frac{1}{4\pi r_\alpha^2 \Delta r} \frac{V}{N^2/2} \langle S_\alpha \rangle$$

Inversion matrix became: $$\frac{\partial \langle S_\alpha \rangle}{\partial V_\gamma} = -\beta\left(\langle S_\alpha S_\gamma \rangle - \langle S_\alpha \rangle \langle S_\gamma \rangle\right)$$

we can reconstruct pair potential from RDF

Additional comments:

- Relationsip " pair potential $\Leftrightarrow$ RDF " is unique
  (Hendeson theorem: R. L. Henderson, Phys. Lett. A 49, 197 (1974)).

- In practice the inverse problem is often ill-defined, that is noticeable different potentials yield RDFs not differing by eye on a graph

- Another scheme to correct the potential (Iterative Boltzmann inversion):
  $V^{(n+1)}(r) = V^{(n)}(r) + kT \ln(g^{(n)}(r)/g^{ref}(r))$

  - may yield different result from IMC though RDFs are similar
  - may not completely converge in multicomponent case
  - may be reasonable to use in the beginning of the IMC iteration process

- Analogy with Renormalization group Monte Carlo
  (R.H.Svendsen, PRL, 42, 859 (1979))

Molecular (multiple-site) systems

1) a set of different site-site potentials
2) intramolecular potentials: bonds, angles, torsions

We use the same expression:

$$H = \sum_{\alpha} V_{\alpha} S_{\alpha}$$

But now α runs over all types of potentials: that is, over all site-site potentials and over all intramolecular potentials, and within each type of potential: over all relevant distances

If α corresponds to a intramolecular potential,
then $\langle S_{\alpha} \rangle$ is the corresponding bond, angle or torsion distribution

Treatment of electrostatics:

If sites are charged, we separate electrostatic part of the potential as:

$$V_{tot}(r_{ij}) = V_{short}(r_{ij}) + \frac{q_i q_j}{4 \pi \varepsilon_0 \varepsilon r_{ij}}$$

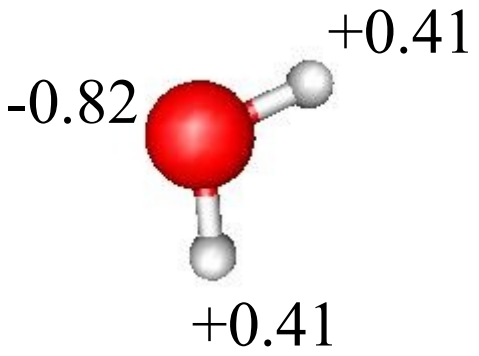
q_i - sum of charges from atomistic model

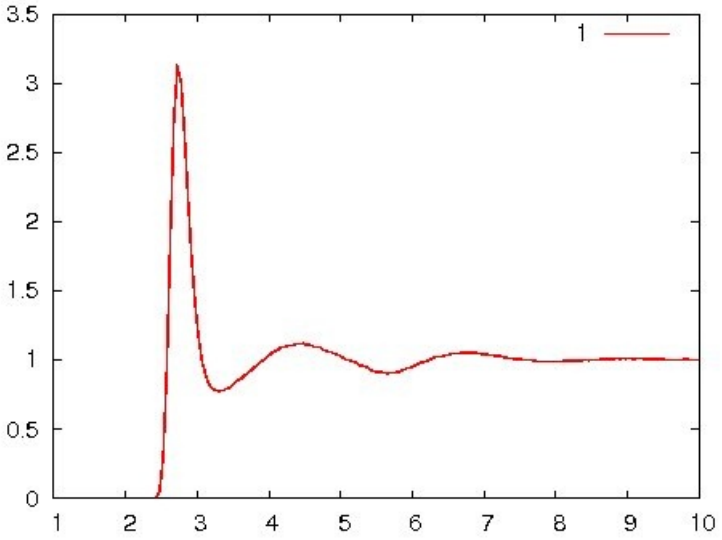
ε - dielectric permittivity: either experimental; or extracted from fitting of asymptotic behaviour of effective potentials)

Electrostatic part: computed by Ewald summation (or PME)

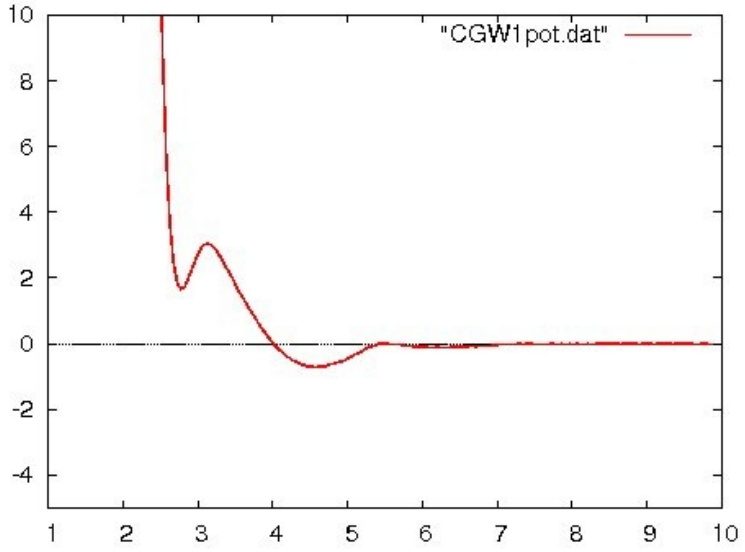
V_{short} - updated in the inverse procedure

Test example: Single site water model





O-O RDF



Potential

Problems: P ≈ 9000 bar ; E ~ + 1kJ/mol - not a liquid state,...

Thermodynamic corrections

We can try to fit some thermodynamics properties (P, E or μ) keeping the potential (approximately) consistent with RDFs

Pressure corrections:

1) add function $\Delta V = A(1\text{-}r/r_{cut})$ (Wang et al, Eur.Phys.J. E, 28, 221 (2009)

2) compute RDF and invert it in two-phase system

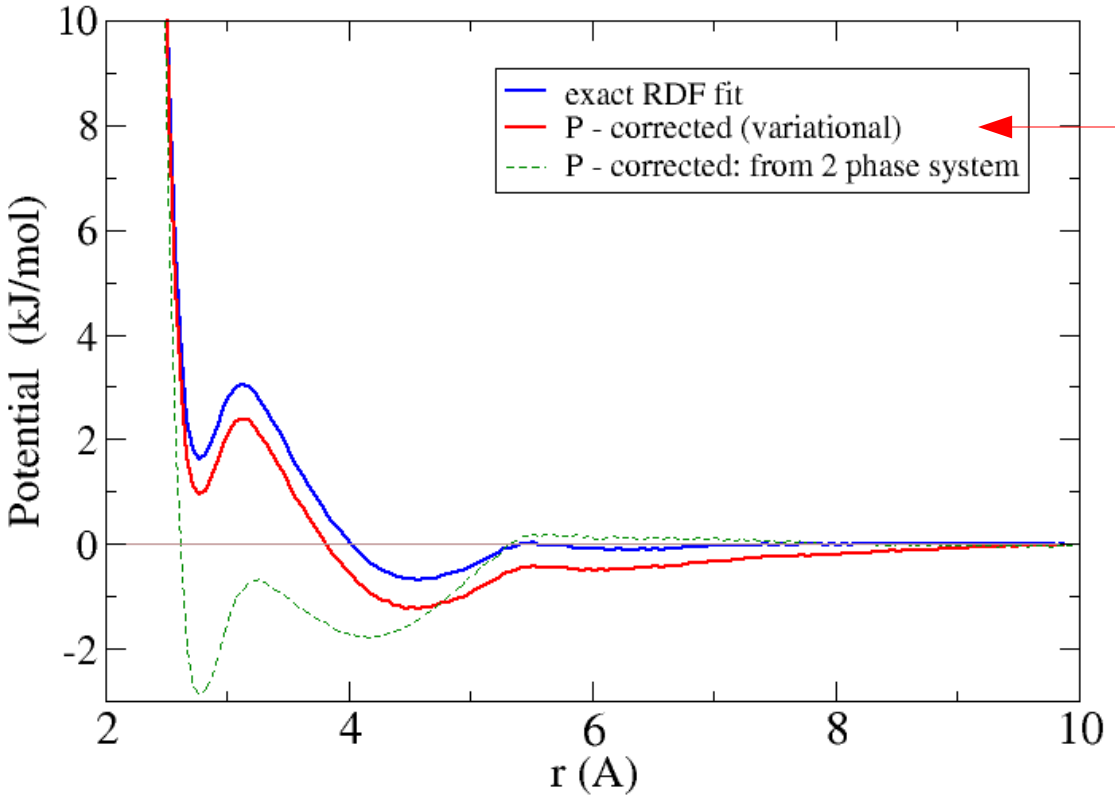
(Lyubartsev et al, Faraday Discussions, 144, 43, (2010))

3) "Variational" inverse MC:

Minimize $\sum_\alpha (\langle S_\alpha \rangle - S_\alpha^*)^2$ under constraints $P = P^*$, $\mu = \mu^*$ ...

(without constraints – equivalent to the standard IMC)

Pressure-corrected 1-site water CG potential



P= 5 +/-10 bar

RDF coincide
within thickness of line

But:

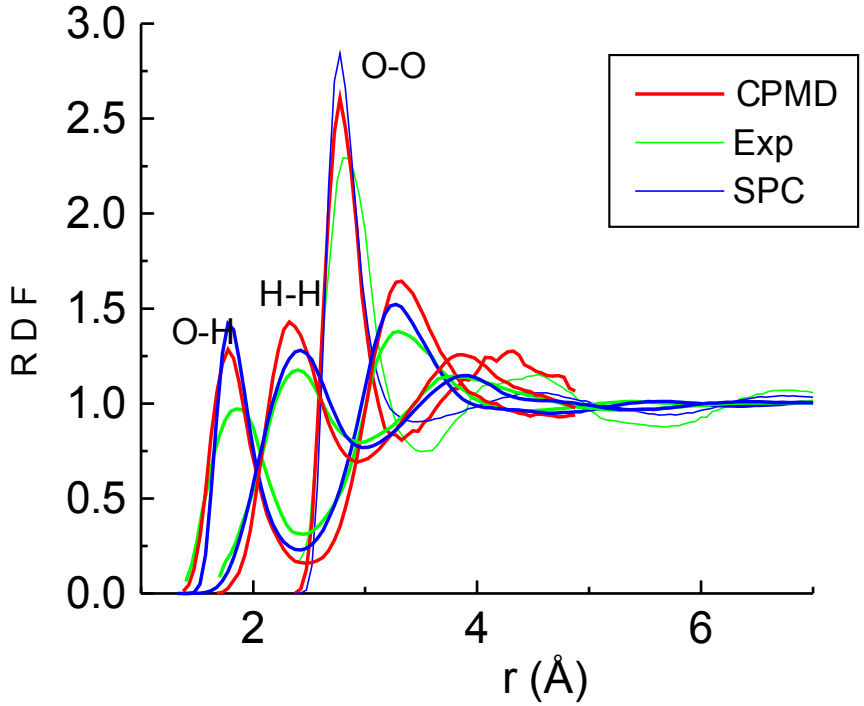
E = -12.5 kJ/mol (exp: -41)

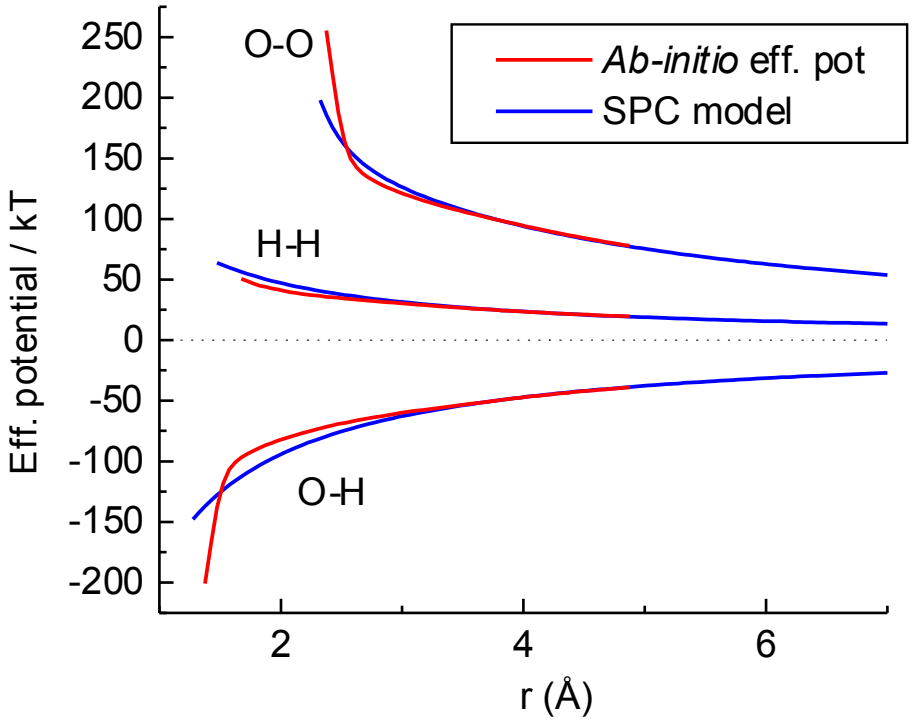
μ = -9.6 kJ/mol (exp: -24)

From ab-initio to atomistic: ab-intio derived 3-site water model

Reference simulation: ab-initio (CPMD) simulation of 32 water

Some details: Box sixe 9.86 Å ,
BLYP functional
Vanderbilt pseudopotentials 25 Ha cutoff
timestep 0.15 fs
25 ps simulation at 300 C





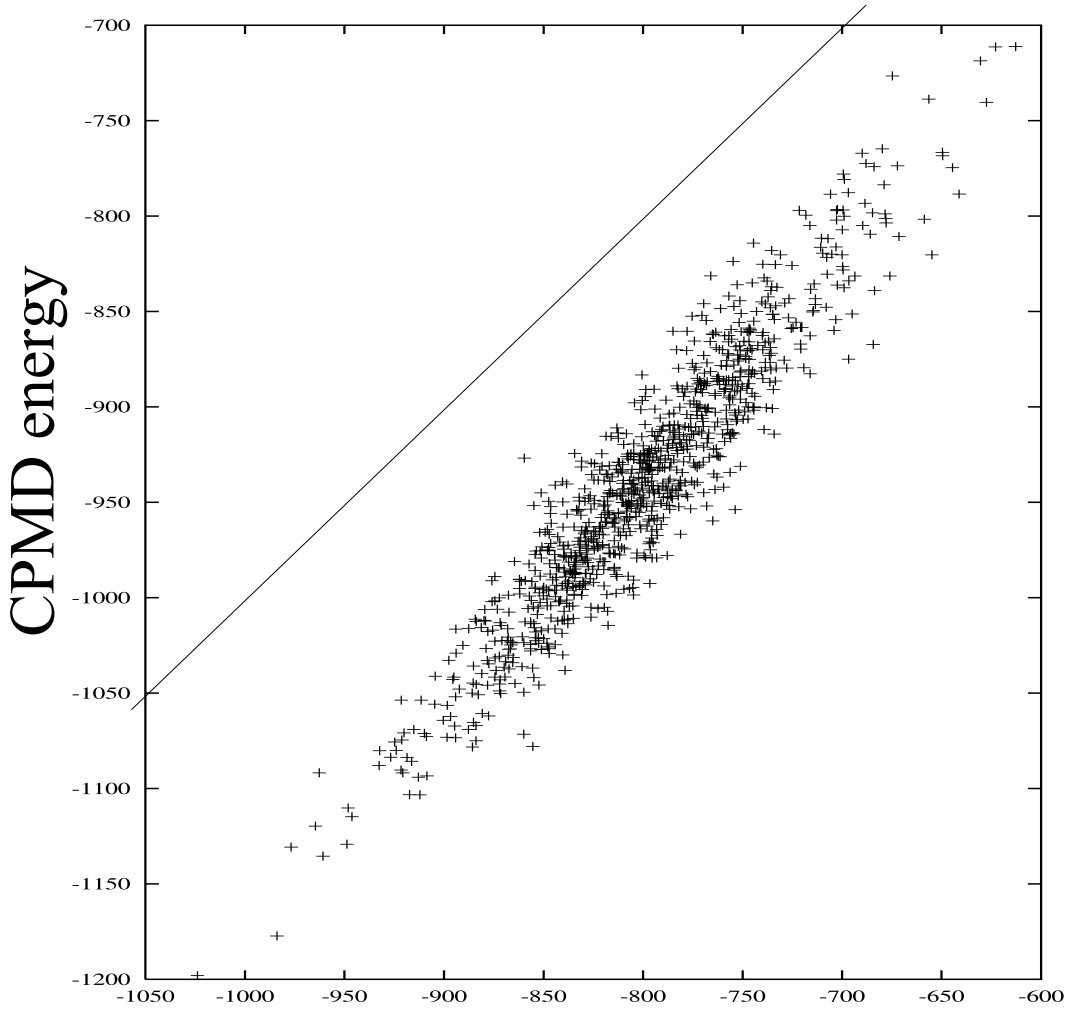
Some insight

Average energy is -25 kJ/mol
Experimental: -41 kJ/mol.

Comparison with CPMD:

$E(32\ H_2O) - 32\ E(H_2O,opt) =$
-29 kJ/mol

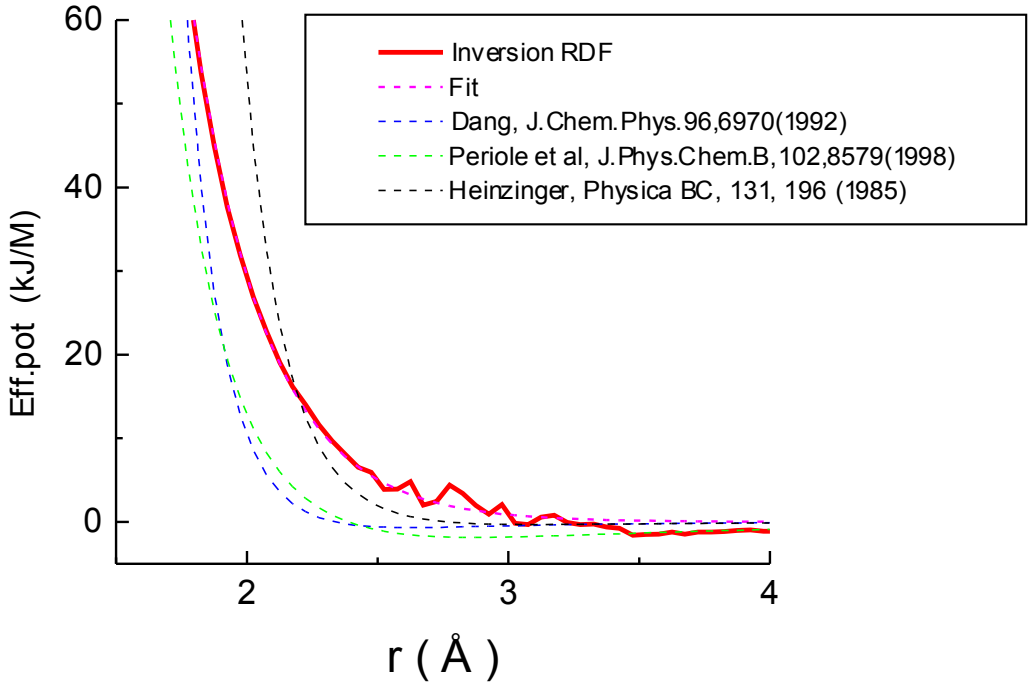
Difference about 4 kJ/mol
due to many body effects



Li^+ - water ab-initio potentials

Input RDF-s obtained from 5+20 ps Car-Parrinello MD for 1 Li+ ion in 32 water molecules (A.P.Lyubartsev, K.Laasonen and A.Laaksonen, J.Chem.Phys., 114,3120,2001)

IMC procedure: SPC model for water; only LiO potential varied to match ab-initio LiO RDF



Non-electrostatic part of the LiO potential may be well fitted by:

$$U_{LiO} = A\ exp(-br)$$

where A = 37380 kJ/M
b = 3.63 $Å^{-1}$

Atomistic simulations: Egorov et al, J.Phys.Chem. B, 107, 3234(2003)

From atomistic to coarse-grained: solvent-mediated potentials.

Atomistic: All-atom simulations (MD) with explicit solvent (water).
State of art: < 100000 atoms - box size ~ 80-100 Å
+ Time scale problem

CG: Coarse-grain solutes; use continuum solvent (McMillan-Mayer level),

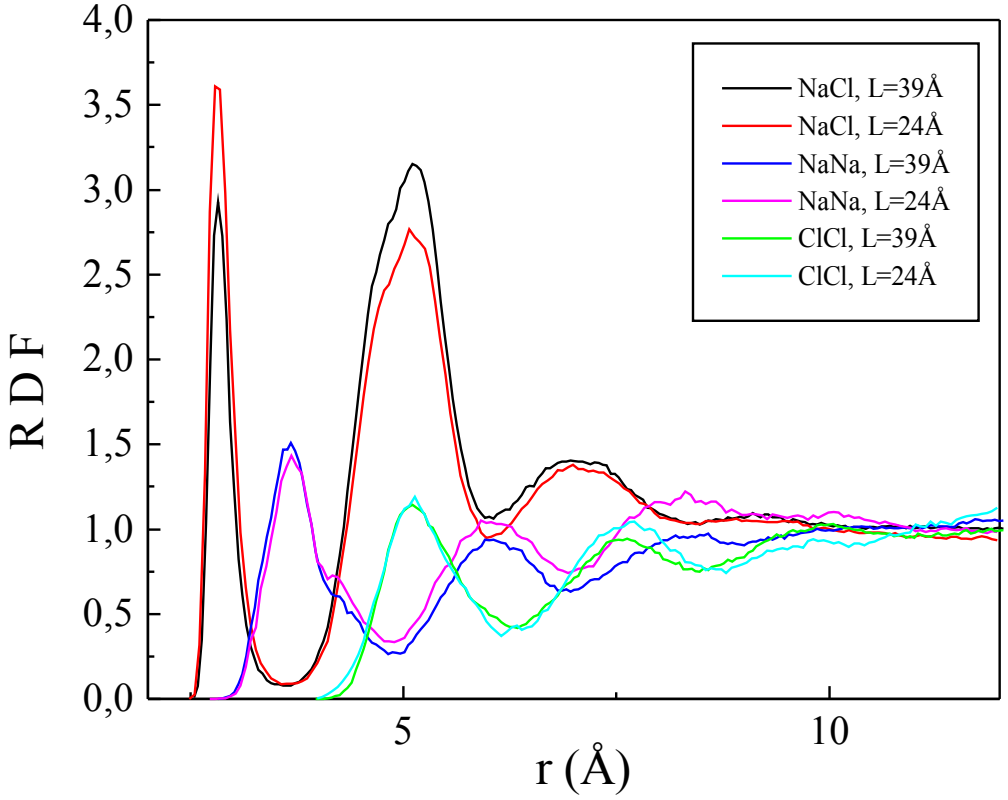
From 60-ties: primitive electrolyte model: ions - hard spheres interacting by Coulombic potential with suitable ε, ion radius - adjustable parameter

We can now build effective solvent-mediated potential, which takes into account molecular structure of the solvent.

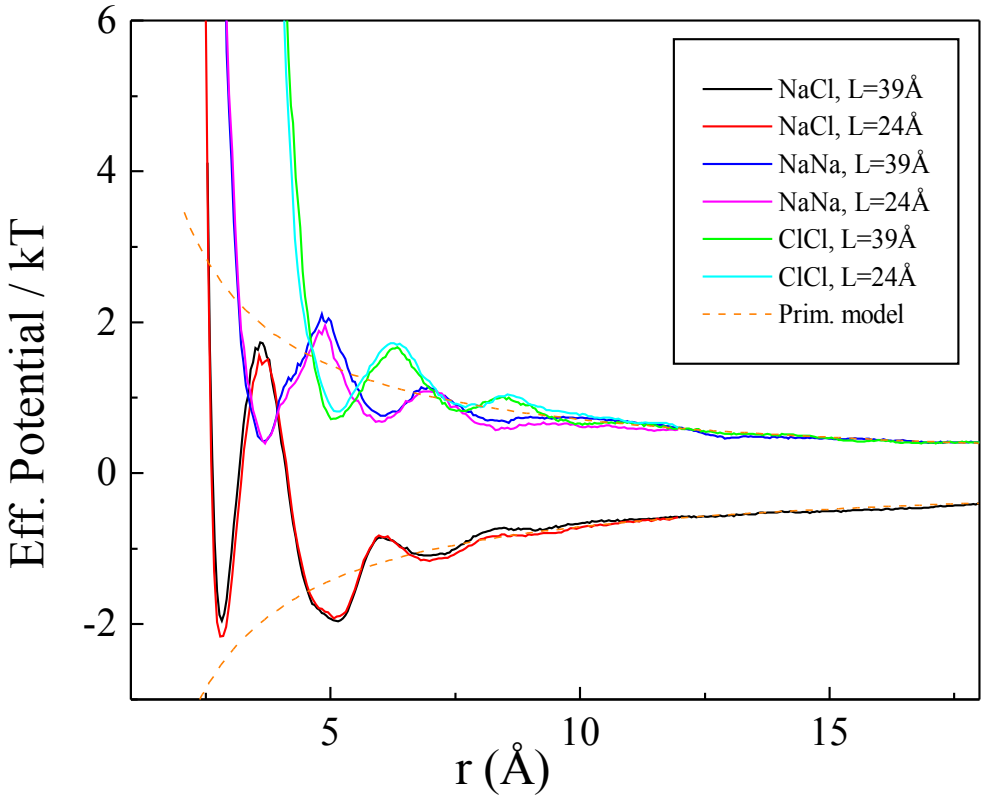
RDF from atomistic MD ⇨ **Implicit solvent effective potentials**

NaCl ion solution

Ion-ion RDFs (from atomistic MD)

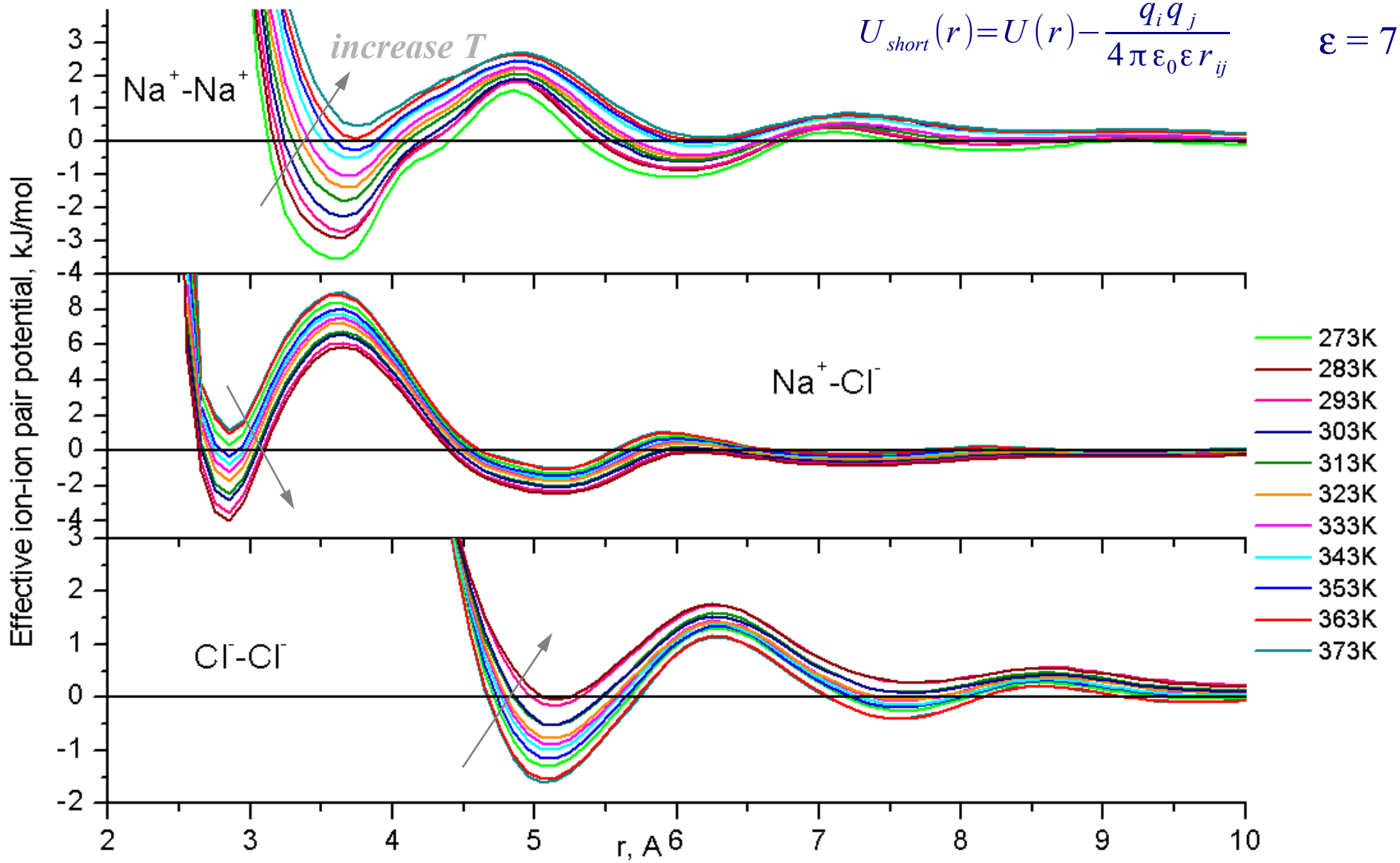


Ion-ion effective potentials

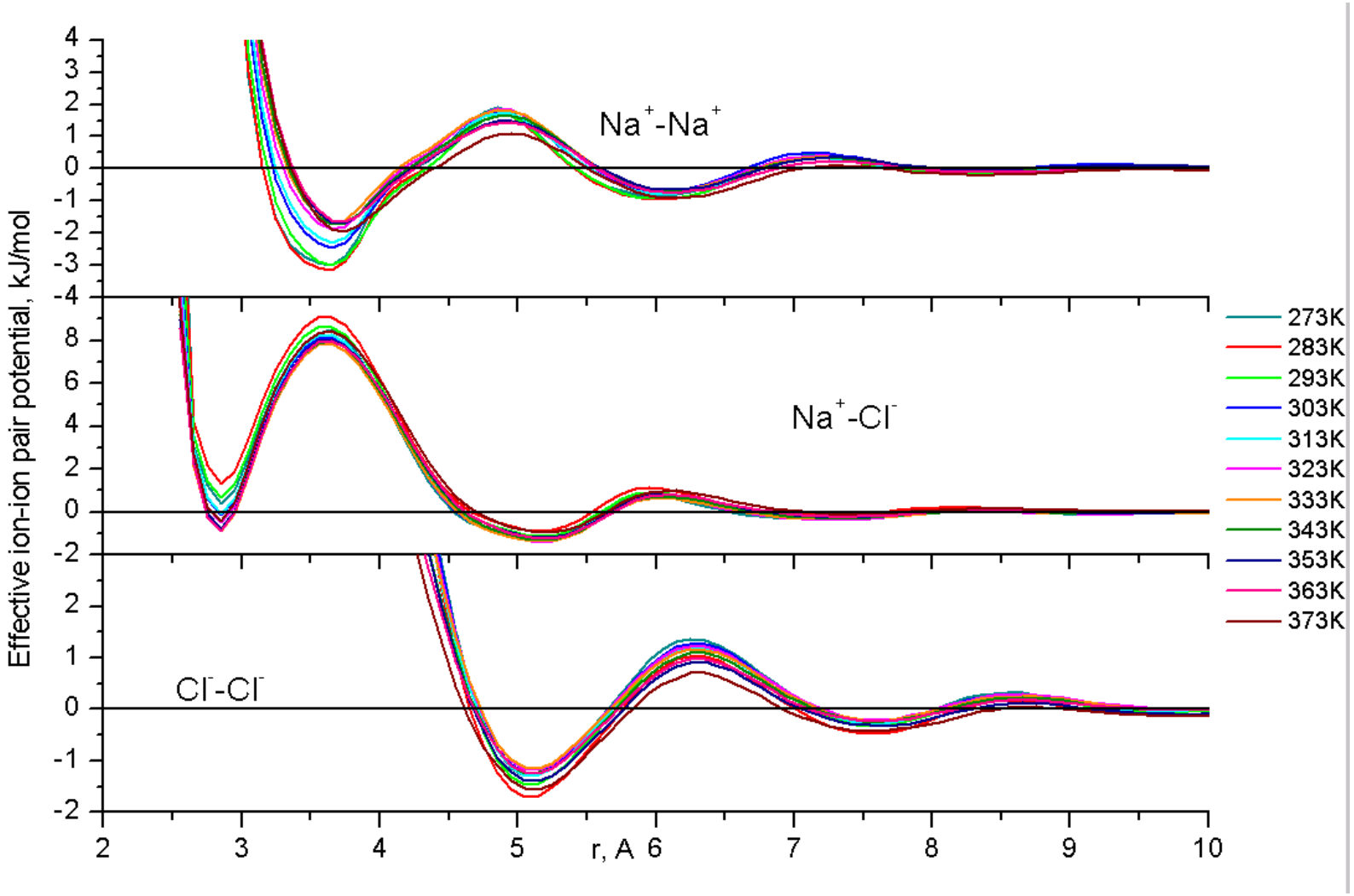


Coulombic asymptotic: $$U(r) = U_{short}(r) + \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon r_{ij}}$$

Temperature dependence



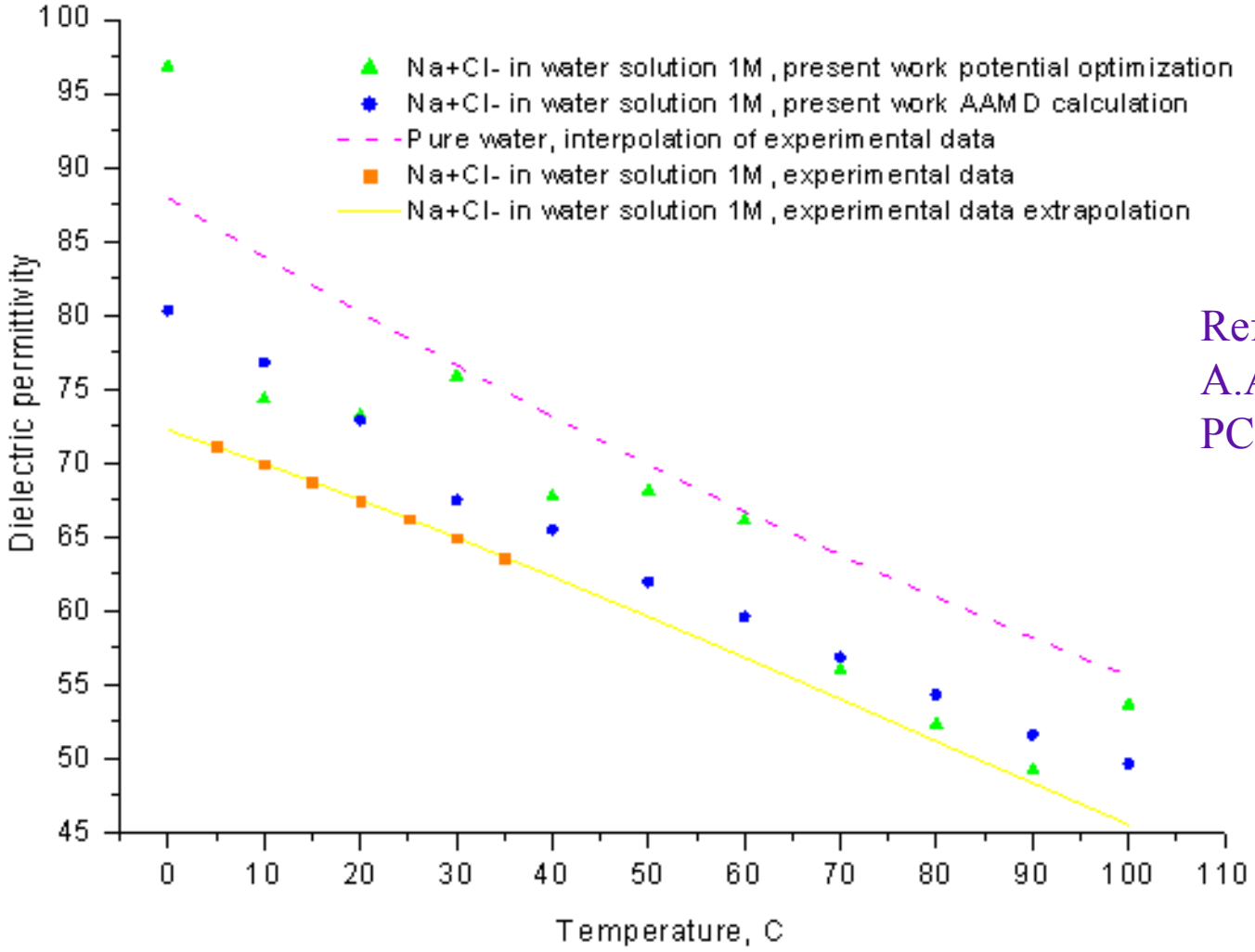
“Short range” effective potential: after optimizing ε



Temperature dependence can be effectively included in *ε*

From tail asymptotic – extract "dielectric permitttivity"

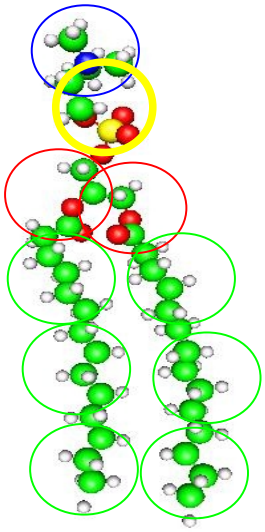
Fit ε to minimize $U_{short}(r) = U(r) - \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon r_{ij}}$ outside some cutoff

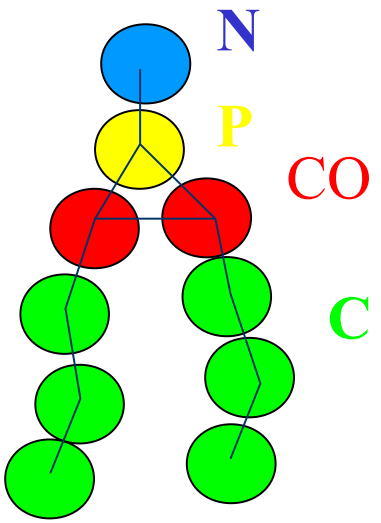


Reference:
A.A.Mirzoev, A.P.Lyubartsev
PCCP 13, 5722 (2011)

Coarse-grain lipid model

DMPC lipid





- 4 different groups -> 10 site-site pairs:
  10 RDFs and eff. intermolecular potentials
- 5 bond potentials: N-P, P-CO, CO – CO, CO-C, C-C)
- 5 angular potentials: N-P-CO; P-CO-C; CO-CO-C,CO-C-C; C-C-C:

In total: 10 site-site intremolecular and
10 intramolecular bond and angle potentials

Earlier work: A.P.Lyubartsev, Eur.Biophys.J, 35,53 (2005) – without angular potentials

All-atomic molecular dynamics

All-atomic MD simulation was carried out:

- 16 lipid molecules (DMPC) dissolved in 1600 waters

+ complementary simulations with 3 other lipid/water ratio

- Initial state - randomly dissolved

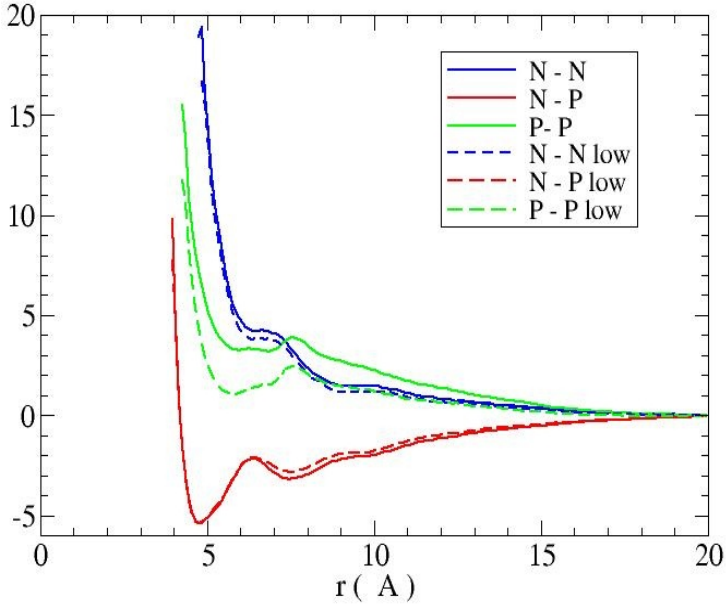
RDFs calculated during 400 ns after 100 ns equilibration

- Force field: CHARMM 27 (modified according to Högberg et al., J.Comput.Chem., 29, 2359(2008)), water - TIP3P

- T=303 K

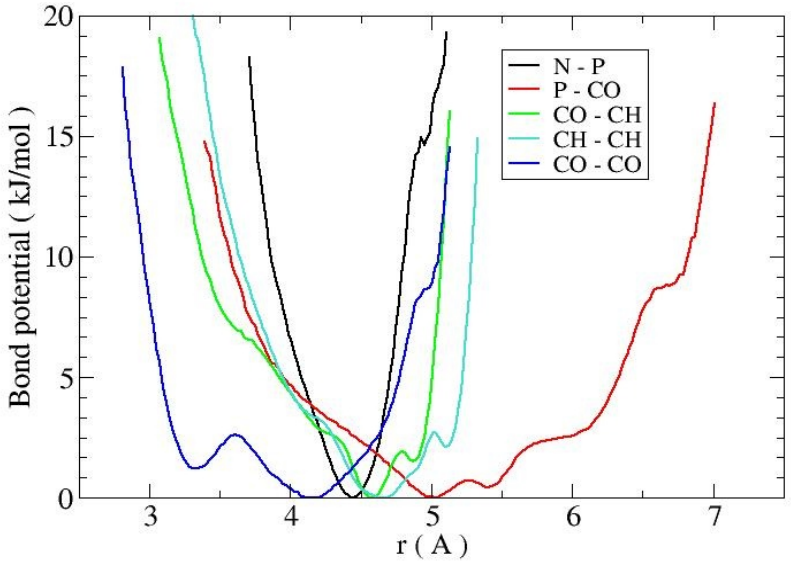
Effective potentials:

Non-bonded potentials

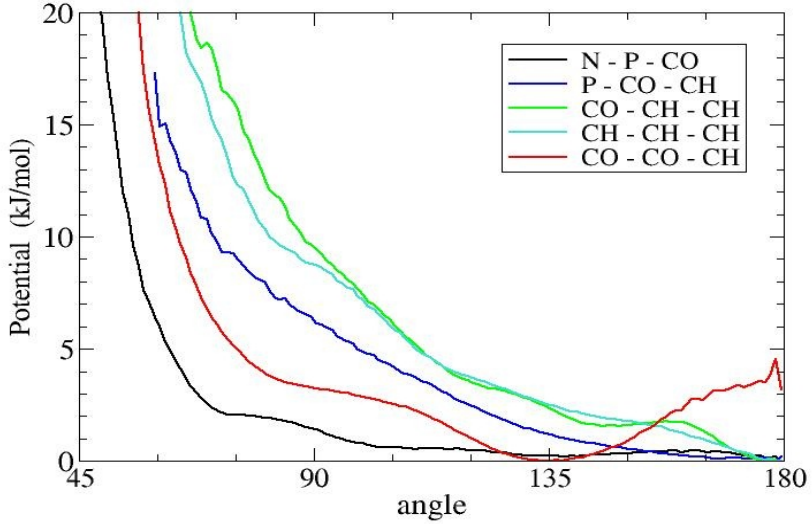


Lipid concentrations in atomistic MD:
a) 50/1500 lipids/H2O
b) 16/1600 lipids/H2O

Bond potentials

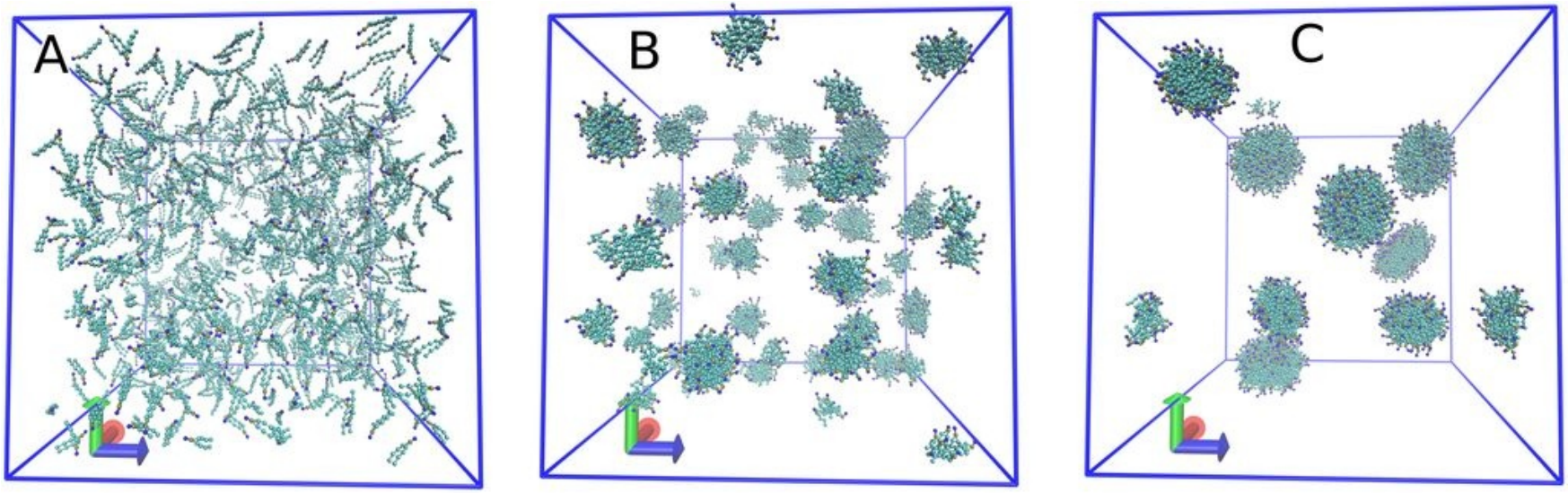


Angle potentials

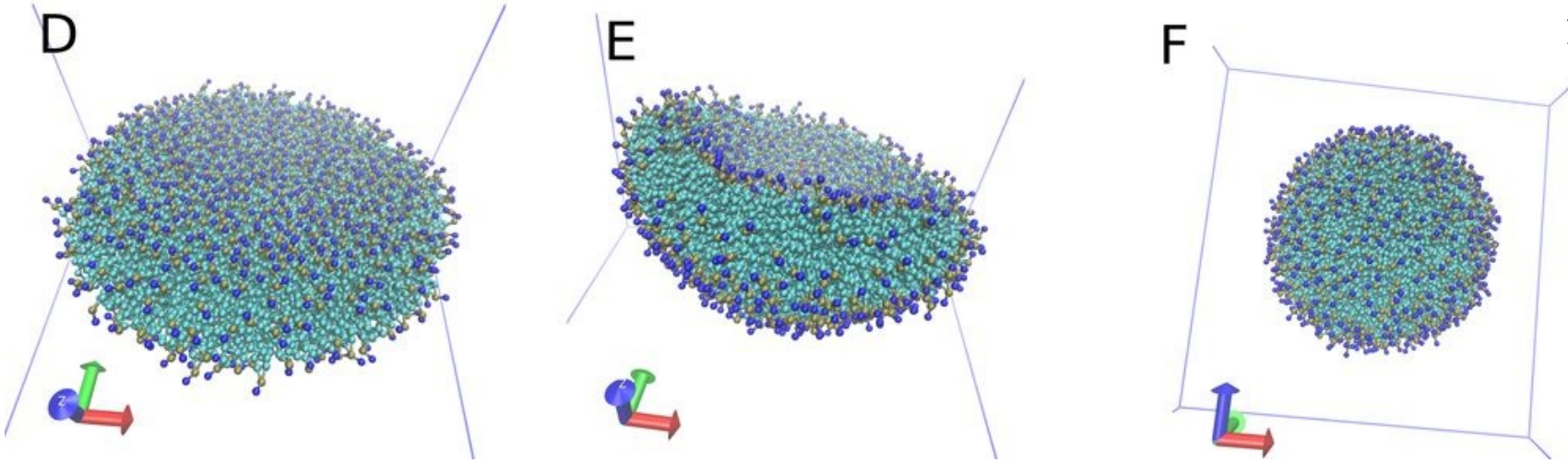


Lipid selfassembly:

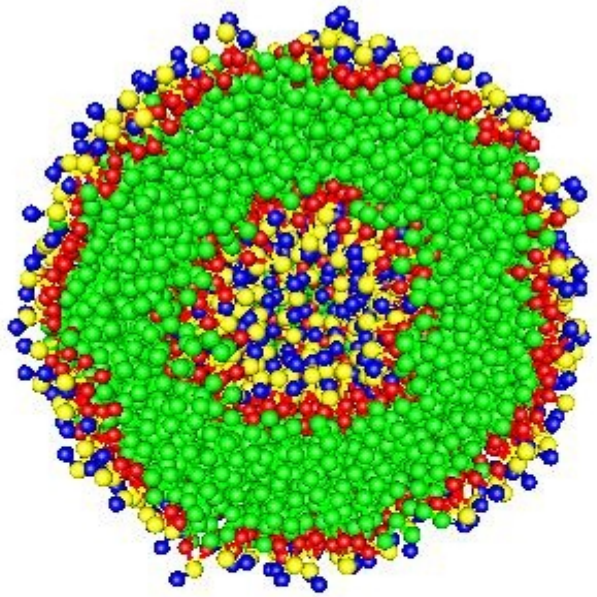
movie 1



movie 2

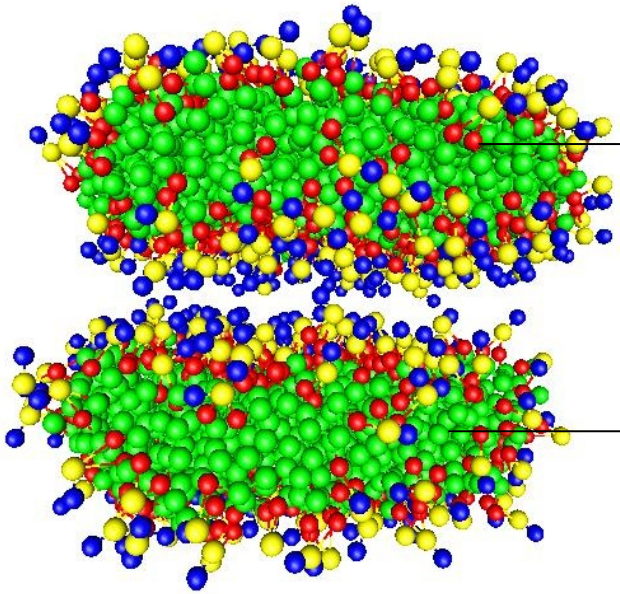


Larger systems (1000 or more lipids) form vesicles:



spherical vesicle
(cut in the middle)
formed by self-assembly
of 1000 DMPC lipids

Smaller systems:
remain as bicells
Sometimes: “double bicell”



Software

Any method can be used only if software is available...

MagiC: Software package implementing Inverse Monte Carlo and Iteraive Boltzmann inversion for calculation of effective potentials for coarse-grained models of arbitrary molecular systems

v 1.0 Released: 21 May 2012

Web page: http://code.google.com/p/magic/

Features:

Interactions:

- tabulated bond and angle potentials
- non-bonded tabulated potentials between different site types
- electrostatic outside cutoff by Ewald

Input (atomistic) trajectories:

- from Gromacs, NAMD, MDynamix, xmol(xyz) format

Methods:

- Iterative Boltzmann inversion and inverse Monte Carlo

Other

- parallel Monte Carlo sampling on many processors
- automatic control of precision
- written in Fortran (2003)

Workflow

1. CGtraj : Reads an atomistic trajectory and creates a CG trajectory according to the given mapping scheme

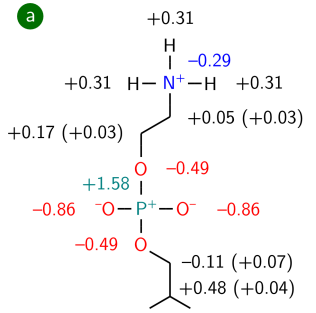
2. RDF: Computes RDFs between all different CG types and bond and angles distributions

3. Magic: Uses IMC or IBI to extract potentials from RDFs and bond and angles distributions. Returns effective potentials.

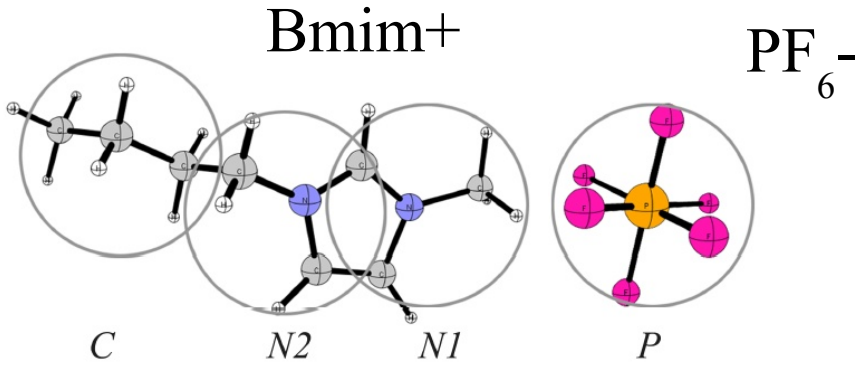
4. Utilities: e.g, convert effective potentials to Gromacs input

A few current applications – in progress

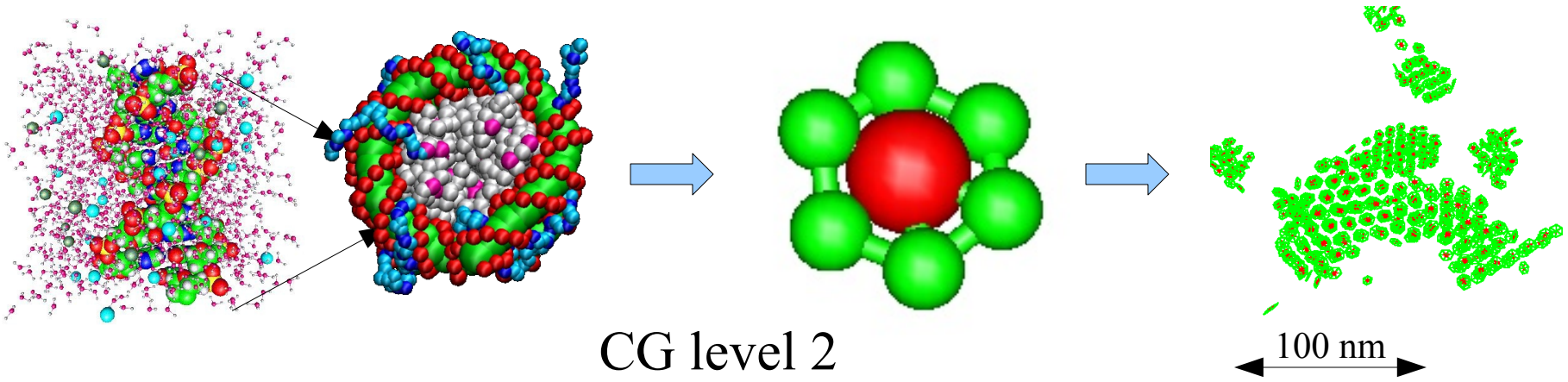
- Other lipids, e.g phosphatydylethanolamine (PE):



- Ionic liquds:



- NCP (nucleosome core particle) selfassembly



Acknowledgements

Coworker and collabolators:

- Alexander Mirzoev
- Joakim Jämbeck
- Andrei Egorov
- Zia ur Rehman
- Aatto Laaksonen
- Aymiric Nayome
- Yonglei Wang
- Lars Nordenskiöld
- Nikolai Korolev
- Yanping Fan

- $$: Swedish Research Council (Vetenskapsrådet)
- CPU: Swedish National Infrastructure for Computing (SNIC)