AN $O(N)$ MULTI-SCALE $N$-BODY APPROACH FOR SIMULATING POLARIZABLE MICROSCOPIC SYSTEMS

Workshop Maths/Industry | Michel Masella | Life Science Division

28 APRIL 2017
Why to use a classical formalism and empirical potentials to model protein systems in solution: efficiency

- Much faster computations compared to quantum approaches (DFT, MP2…)
- Large systems (thousands up to millions of atoms)
- Long simulations (from the ns up to the μs scale)

Example: simulating the mature HIV-1 capsid structure

128 000 cores of the Cray “Blue Waters”, 300 ns, about $60 \times 10^6$ atoms ($50 \times 10^6 = \text{water}$)

Anton, a special-purpose machine for molecular dynamics simulation
Shaw et al, ACM SIGARCH, 35 (2007) 1-12

How fast folding proteins fold
The main drawback of MM approach: accuracy

Common force-fields are based on additive potentials (CHARMM, AMBER, OPLS …)

1 - The model parameters are constant along an MD trajectory

2 – Reliability of such an approach when mimicking microscopic electrostatic interactions

\[ U_{\text{Coulombic}} = \sum_{i<j} \frac{\delta_i \delta_j}{r_{ij}} \]

\( \delta_i, \delta_j \) are static charges

3 – Experimental results are commonly used for assigning/refining parameters, transferability?

4 - Ok, you may say that for a homogeneous and isotropic system, you are using a mean field approach...

However, what about interfaces?
These phenomena cannot be accounted for using static charges to model the molecular electronic cloud properties.
They allow to account for environmental effects on electrostatic molecular properties considered in classical molecular dynamics

Many approaches have been proposed:

- Fluctuating charge approaches: the “static” $\delta_i$ charges are allowed to fluctuate
- Drude oscillators: “extra” atomic charges are introduced
- Induced dipole moments: new degrees of freedom are introduced

$$
\mu_i = \alpha_i \left( E_i - \sum_{i=1, i \neq j}^{N} T |r_j - r_i| \mu_j \right) \Rightarrow \text{to be solved iteratively, } O(N^2)
$$

$T$ is the dipolar tensor and $\alpha_i$ is the center $i$ polarizability, usually isotropic
$E_i$ is the electric field acting on center $i$
$\mu_i$ is the induced dipole moment on center $i$
Accounting for inter atomic interactions with a weak covalent character

\[ U_{ct} = \sum_{l \in \text{water}} D_e^c \exp(-\beta \cdot r_{Th-l}) \]

\[ D_e^c = D_e \left[ 1 - \xi \sum_{i \in \text{water}, i \neq l} \exp \left( -\left( r_{Th-i} - r_e \right)^2 / \gamma_r \right) \right] \]

Seems computationally demanding, however scales as \( O(N) \)
Benchmark binding energies of ammonium and alkyl-ammonium ions interacting with water. Are ammonium-water hydrogen bonds strong?

V. Vallet and M. Masella, CPL, 618 (2015) 168
AB INITIO FORCE-FIELDS: LIQUID WATER

![Graphs showing enthalpy, density, and dielectric constant vs. temperature for liquid water at different pressures.](image-url)
Water critical point

\[ \rho_\pm = \rho_c + b_0 |T - T_c| \pm (b_1 |T - T_c|^{\beta_1} + b_2 |T - T_c|^{\beta_1+1/2}) , \]

Water surface tension

\[ \gamma(T) = a_\gamma (T_c - T)^{5/4} + b_\gamma (T_c - T)^{9/4} , \]

Real et al, J Chem Phys, 2013
\[ \Delta H_{\text{solv}}(\text{H}^+) = \Delta H_{g \rightarrow aq}(\text{PH}) - \Delta H_{g \rightarrow aq}(\text{P}) + \Delta H_{g}^{\text{prot}}(\text{P}) + \Delta H_{aq,\text{dissociation}}^0(\text{PH}) \]
A multiscale solvent coarse grained approach
A SOLVENT COARSE GRAINED APPROACH

Explicit solvent approach, need of FFT!

Coarse grained approach
Efficiency and microscopic solvation properties

$\alpha_S = \frac{1}{4\pi} \left( 1 - \frac{1}{\varepsilon_s} \right) \times \frac{1}{\rho_s}$

\textit{i.e.} $\alpha_S \propto \Delta v_s$

$F_P[P] = \frac{1}{2} \int \frac{\varepsilon(r)P(r)^2}{\chi(r)} d\nu - \int P(r) \cdot E_S(r) d\nu$

$P(r) = \frac{\chi(r)}{\varepsilon(r)} E_S(r)$

$P_s^{\Delta\nu} = \frac{\Delta \nu P(r)}{\rho_s}$

Implicit solvent approaches
Poisson-Boltzmann
Warshell’s grid approach

In the case of water, number of solvent atoms divided by 3

An explicit water molecule → A nonpolar polarizable particule = a water molecule

What about the liquid water hydrogen bond network?

Systematic truncation of long tail electrostatic interactions

with $R_c = 12$ Å, the solvent cost represents 10 to 30% of the total cpu time

What about electrostatic long tail effects?
Free energy cost $\Delta G$ for creating a cavity within liquid water (Monte-Carlo SPC/E simulations)


Computing $\gamma$ with VdW potentials: large cutoff (15 Å) for estimating it within 10%

HYDROPHOBIC EFFECTS

\[ U_{pp}^{\text{density}} = \sum_{k \leq i \leq N_s} \varepsilon_s^0 \left( n_i^0 - \overline{n}^0 \right)^2 + \varepsilon_s^1 \left( n_i^1 - \overline{n}^1 \right)^2 \]

\( n_i \): solvent local density at the vicinity of a particle

\( 0 \rightarrow \text{1st shell} ; 1 \rightarrow \text{2nd shell} \)

\( \rightarrow \) solvent/solvent interaction truncation : 7 Å


PMF of the methane dimer (300 K, 1 atm)

Methane : solvent structure


Agreeing with the theory of the hydrophobic effect …


...and all atom simulations
Particle polarizability proportional to their individual volume: \( \alpha_s \propto \Delta V_s \)

→ a multi-scale coarse grained approach to handle long tail bulk electrostatic

An explicit water molecule  
A polarizable pseudo-particule = a water molecule  
A larger polarizable pseudo-particule = a volume element of liquid water

\[ Pseudo-particule = a \text{ water molecule} \]
\[ Pseudo-particule = \text{volume element made of 8 water molecules} \]
\[ Pseudo-particule = \text{volume element made of 64 water molecules} \]

\[ Level \ 0 \quad R \approx 12 \text{ Å} \quad Level \ 1 \quad R \approx 30-40 \text{ Å} \quad Level \ 2 \quad Level \ n \]
Now, the bottleneck is handling the interactions within the solute…
Potential of mean force of Cl⁻/Cl⁻ and the coarse grained approach level

(1ns simulation, solvent boxes made of 4096 particles, umbrella sampling, etc…)

\[ \frac{q_1q_2}{4\pi\varepsilon_0 R_{12}} \]

Polarizable Pseudo Particles ⇒ Macroscopic laws of electrostatic

Energetic of creating a cavity within liquid water ⇒ Hydrophobic effects

Polarizable force-field

Supported by the Exascale Computing Research Laboratory

All atom simulations+ PME summation techniques

A water box of 1000 molecules ➔ 3.0 ns per day (1 cpu)
4.5 ns per day (2 cpu)

Coarse grained simulations + solute interactions $O(N^2)$

6000 particles + solute = 900 atoms ➔ 15 ns per day (level 0 + 1 cpu)
13 ns per day (level 1 + 1 cpu)
11.5 ns per day (level 2 + 1 cpu)
9.5 ns per day (level 3 + 1 cpu)

Largest system simulated:

8400 atoms solute + 51 000 particles, 850 ps per day (level 0 + 1 cpu)

➔ Fast Multipole Method !!!

Reference : Intel Sandy-Bridge octo-cores 2.7 Ghz (CURIE)
10x20 ns simulations of three small solvated proteins (240 SB cores/36 h)


No parameter assigned to reproduce protein structures
An $O(N)$ multi-scale $N$-body approach for simulating polarizable microscopic systems
A O(N) FFM APPROACH

\[ \phi(x_b - x_a) = \frac{q_a}{|x_b - x_a|} \]

\[ \phi(x_b - x_a) \approx \sum_{|n| \leq p} \sum_{|m| \leq p - |n|} \frac{(-1)^{|n|}}{n!m!} r_b^n r_a^m \nabla^{n+m} \phi(z_B - z_A) \]

\[ + \mu_a = 2q_q \delta_l_a \]

+ atoms organized via a kd-tree spatial decomposition

Mean error on the forces

Calendélectrostatique avec la FMM

0.0M 1.0M 2.0M 3.0M 4.0M 5.0M 6.0M 7.0M

0 50 100 150 200 250 300 350

Temps normalisé

No. Atomes

0 1 2 3 4 5 6 7
Dipoles solved iteratively; MTS (2f/4fs); black: CURIE@CEA (node = 16 cores); blue: OCCIGEN@CINES (1 node 24 cores)
Mosaic Panicum virus capsid
0.6 M atoms + 1.8 M coarse grained solvent particles
Equivalent to a 6.5 M atom systems

Method scalability
Hazelhen supercomputing system
(HLRS, Stuttgart, Germany)