DE LA RECHERCHE À L'INDUSTRIE

cea

TOWARDS SIMULATING THE CELL MACHINERY AT THE ATOMIC SCALE

"Calculs Parallèles et Applications" | Michel Masella | Life Science Division

www.cea.fr

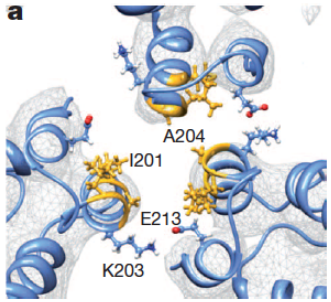
30 SEPTEMBRE 2014

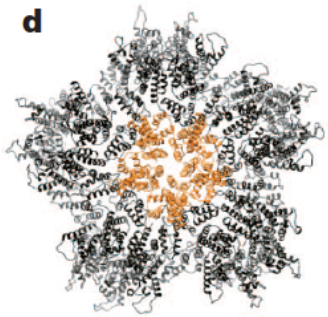
MOLECULAR MODELLING VS QUANTUM CHEMISTRY

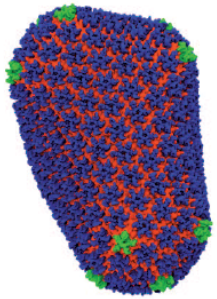
Why to use a classical formalism and empirical potentials to model protein systems in solution: efficiency

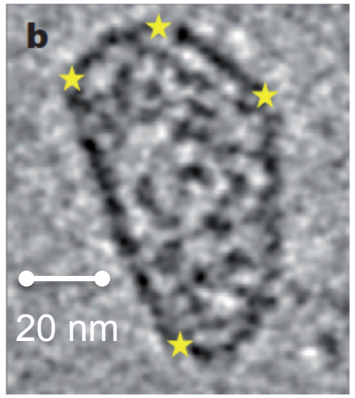
- Much faster computations compared to quantum approaches (DFT, MP2…)
- Large systems (thousands up to millions of atoms)
- Long simulations (from the ns up to the µs scale)

Example : simulating the mature HIV-1 capsid structure









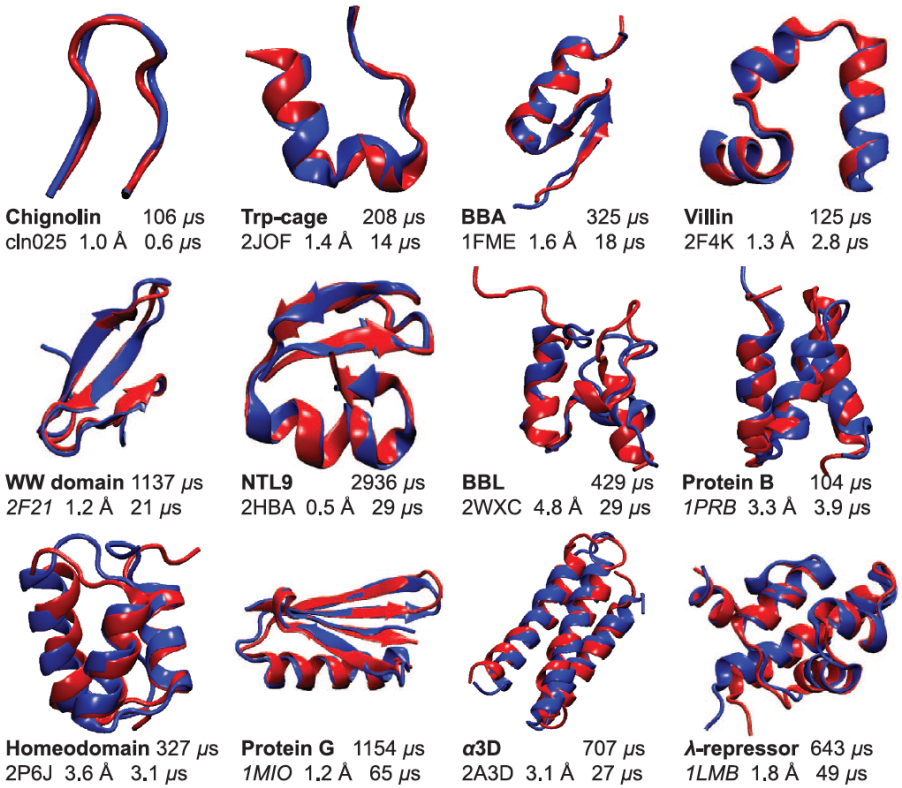
128 000 cores of the Cray “Blue Waters”, 300 ns, about 60.10^6 atoms (50.10^6 = water)

*Zhao et al, Nature, **497** (2013) 643*

THE ANTON MACHINE

Anton, a special-purpose machine for molecular dynamics simulation

Shaw et al, ACM SIGARCH, 35 (2007) 1-12



How fast folding proteins fold

*Lindorff-Larssen et al, Science, **334** (2011) 517*

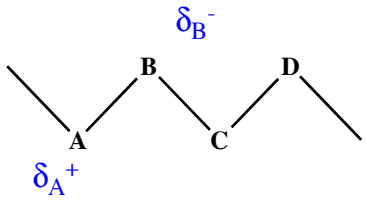
STANDARD MM APPROACHES: ACCURACY ?

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



➔ $U_{Coulombic} = \sum_{i<j} \frac{\delta_i \delta_j}{r_{ij}}$ ➔ δ_i are static charges

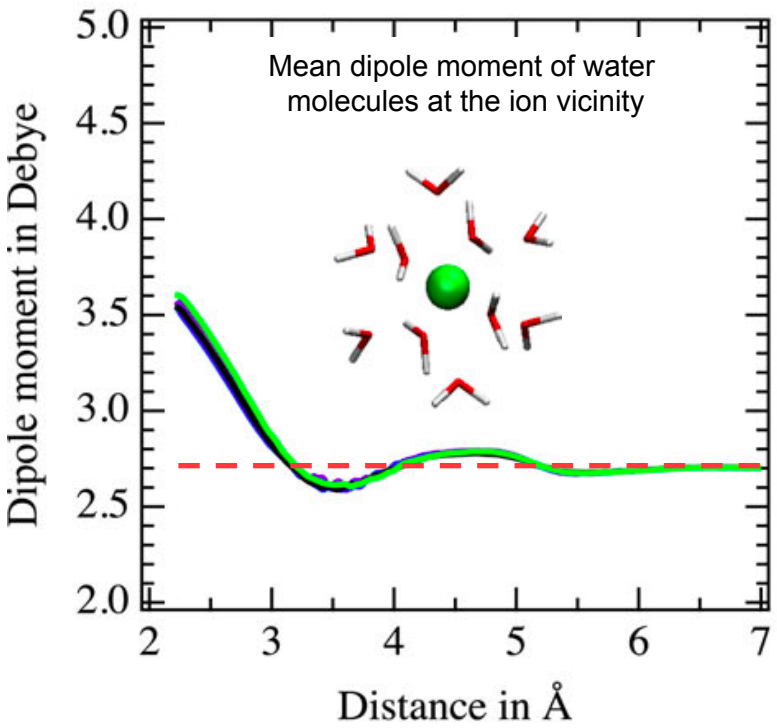
3 – Experimental results are commonly used for assigning/refining parameters, transferability ?

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

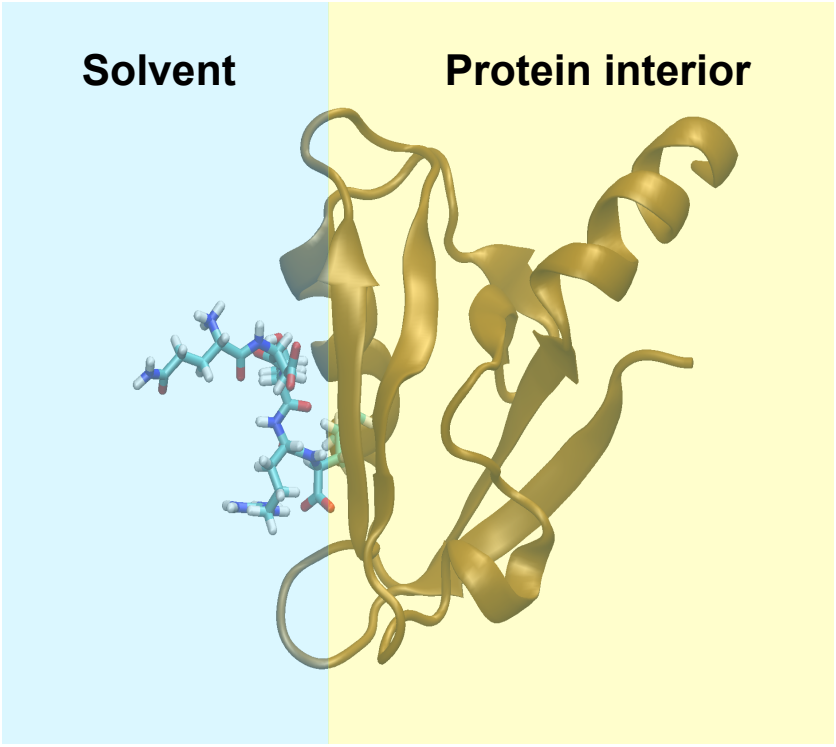
However, what about interfaces ?

POLARIZATION EFFECTS AND INTERFACES

The case of Cm(III) interacting with water



All protein/ligand interactions correspond to an interface problem



These phenomena can not be accounted for using static charges to model the molecular electronic cloud properties

POLARIZABLE FORCE-FIELDS

They allow to account for environmental effects on electrostatic molecular properties considered in classical molecular dynamics

Many approaches have been proposed :

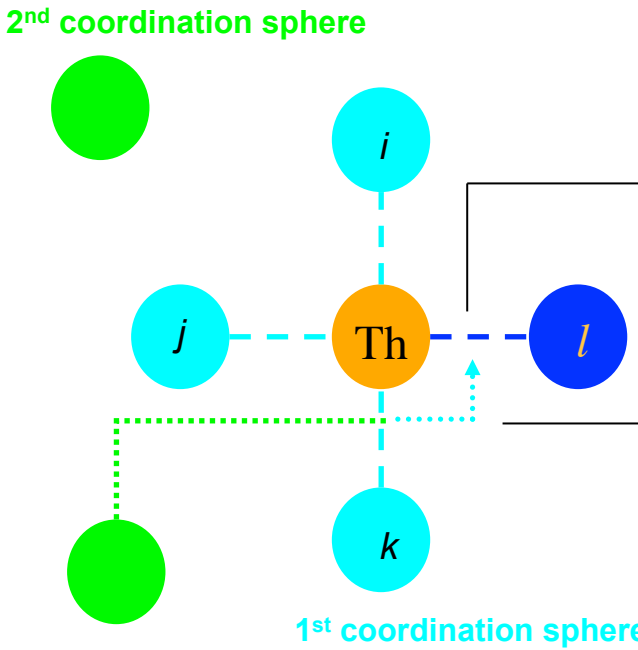
- Fluctuating charge approaches: the “static” δ_i charges are allowed to fluctuate
- Drude oscillators: “extra” atomic charges are introduced
- **Induced dipole moments: new degrees of freedom are introduced**

$$\boldsymbol{\mu}_i = \alpha_i \left(\boldsymbol{E}_i - \sum_{i=1, i \neq j}^{N} \mathbf{T}\left|\boldsymbol{r}_j - \boldsymbol{r}_i\right| \boldsymbol{\mu}_j \right) \Rightarrow \text{ to be solved iteratively, } O(N^2)$$

T is the dipolar tensor and α_i is the center *i* polarizability, usually isotropic
$\mathbf{E}_i$ is the electric field acting on center *i*
$\boldsymbol{\mu}_i$ is the induced dipole moment on center *i*

BEYOND POLARIZATION : CHARGE TRANSFER

Accounting for inter atomic interactions with a weak covalent character



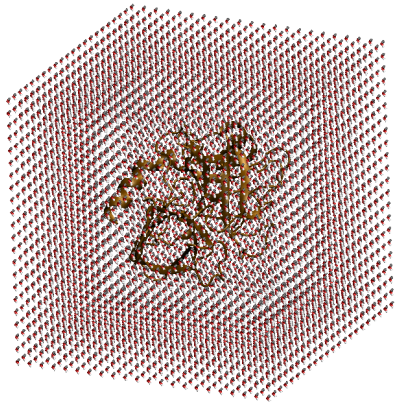
$$U_{ct} = \sum_{l \in water} D_e^c \exp(-\beta \cdot r_{Th-l})$$

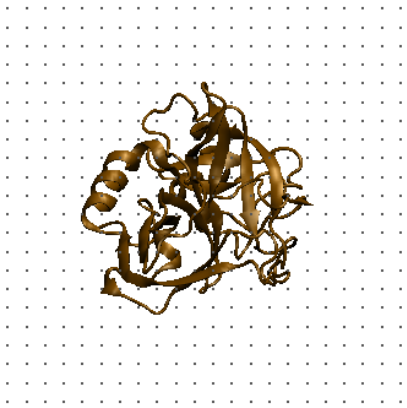
$$D_e^c = D_e \left[1 - \xi \sum_{i \in water, i \neq l} \exp\left(-(r_{Th-i} - r_e)^2 / \gamma_r \right) \right]$$

Seems computationally demanding, however scales as *O(N)*

A SOLVENT COARSE GRAINED APPROACH

Explicit solvent approach, need of FFT !





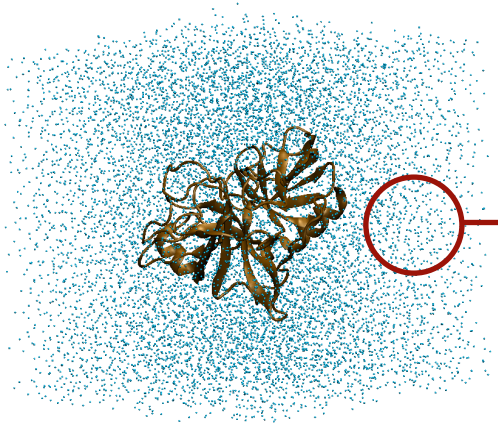
Implicit solvent approaches

Poisson-Boltzman

Warshell's grid approach

Coarse grained approach

Efficiency and microscopic solvation properties



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

i.e. $\alpha_s \propto \Delta v_s$

$$F_P[\mathbf{P}] = \frac{1}{2}\int \frac{\varepsilon(\mathbf{r})\mathbf{P}(\mathbf{r})^2}{\chi(\mathbf{r})} dv - \int \mathbf{P}(\mathbf{r}) \cdot \mathbf{E}_S(\mathbf{r}) dv$$

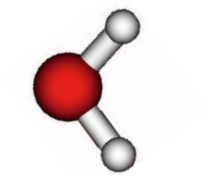
$$\mathbf{P}(\mathbf{r}) = \frac{\chi(\mathbf{r})}{\varepsilon(\mathbf{r})}\mathbf{E}_S(\mathbf{r}) \longrightarrow \mathbf{p}_k^s = \Delta v \mathbf{P}(\mathbf{r}) = \frac{\mathbf{P}(\mathbf{r})}{\rho_s}$$

Haduong et al, J. Chem. Phys., ***117*** *(2002) 541*

Masella et al, J. Comput. Chem., ***29*** *(2008) 1707*

ORIGIN OF COARSE-GRAINED APPROACH EFFICIENCY

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



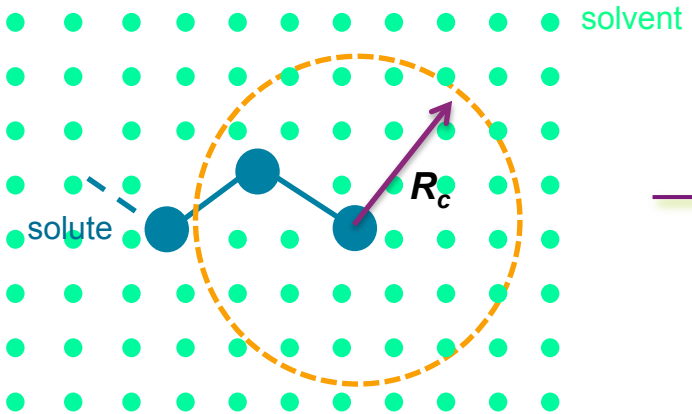
An explicit water molecule

A nonpolar polarizable particule
= a water molecule

What about the liquid water hydrogen bond network ?

A solution from DPD

Systematic truncation of long tail electrostatic interactions



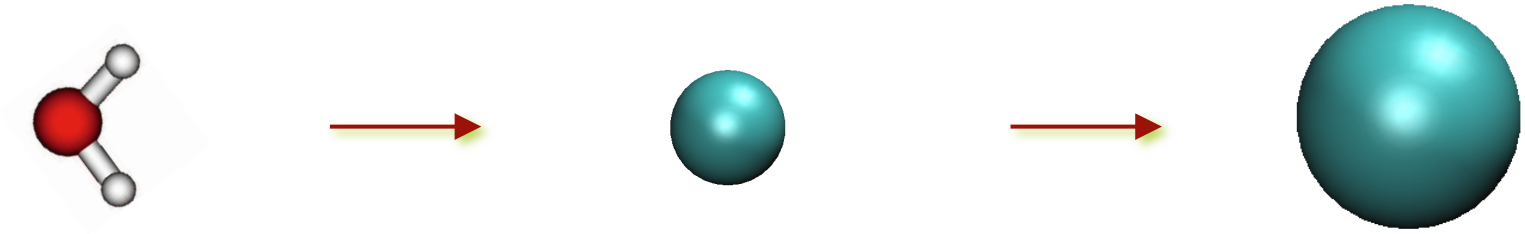
What about electrostatic long tail effects ?

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

« MULTI-SCALING » THE COARSE GRAINED APPROACH

Particle polarizability proportional to their individual volume : $\alpha_s \propto \Delta v_s$

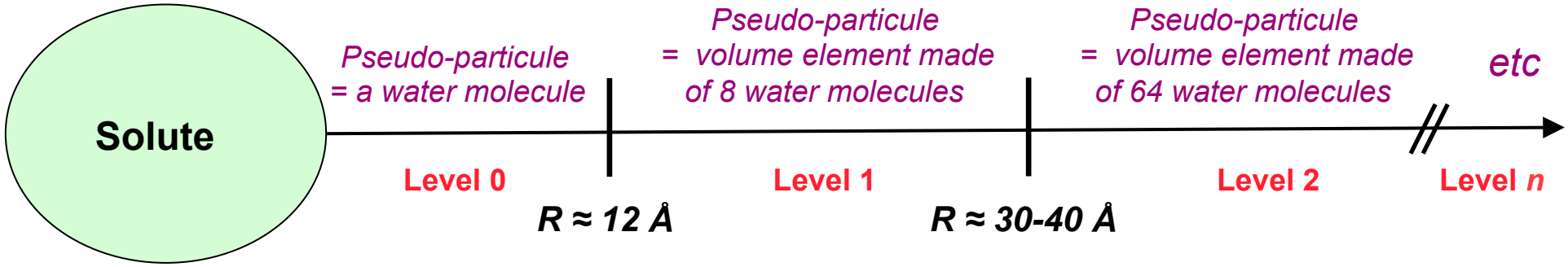
➔ **a multi-scale coarse grained approach to handle long tail bulk electrostatic**



An explicit water molecule

A polarizable pseudo-particule
= a water molecule

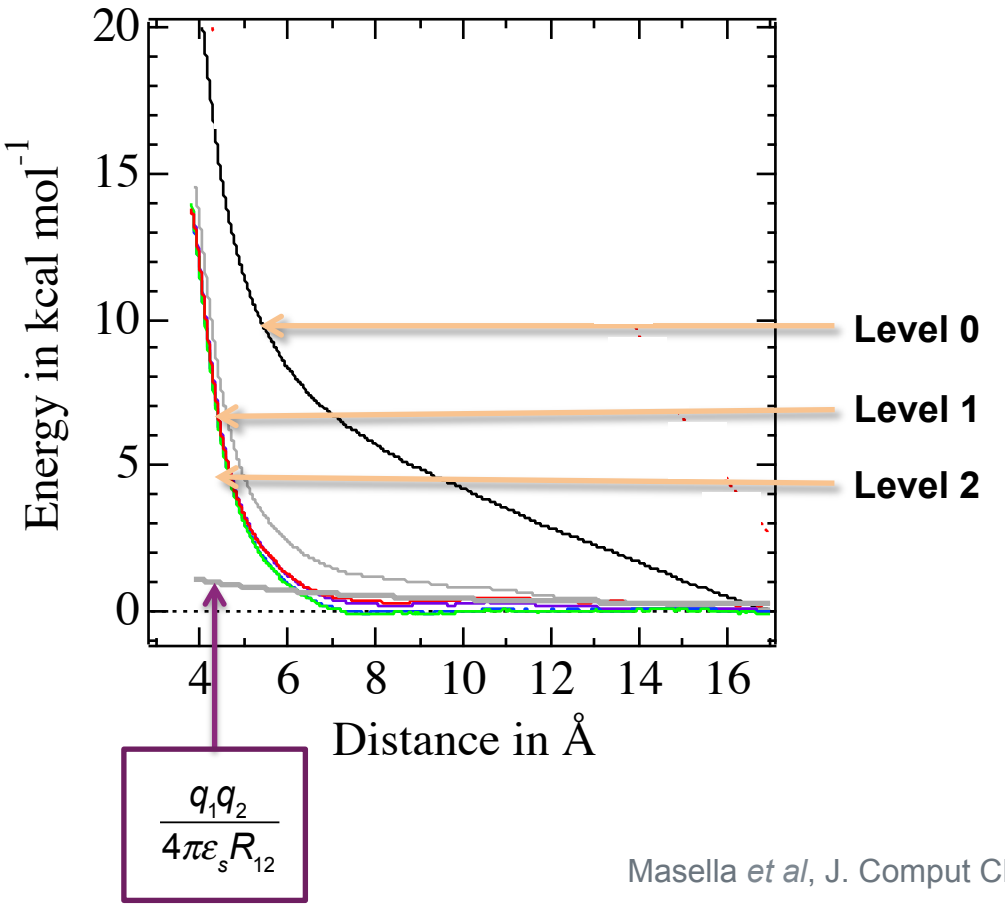
A larger polarizable pseudo-particule
= a volume element of liquid water



« MULTI-SCALING » THE COARSE GRAINED APPROACH

Potential of mean force of Cl^-/Cl^- and the coarse grained approach level

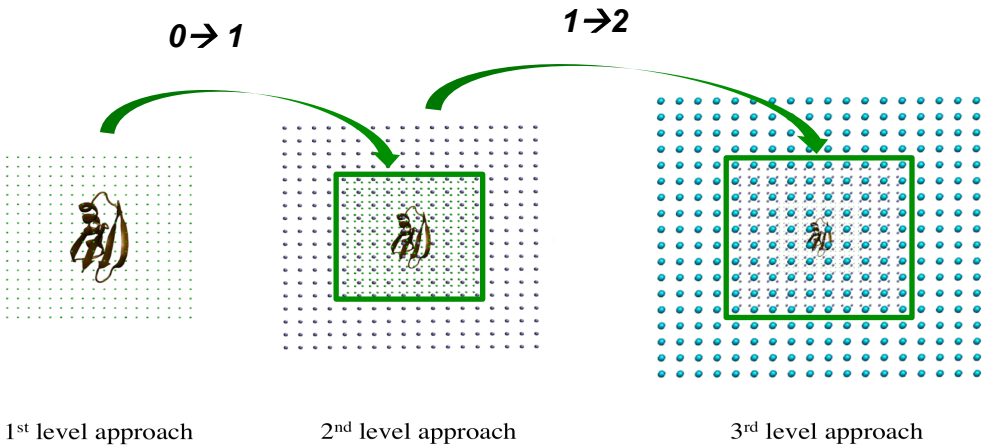
(1ns simulation, solvent boxes made of 4096 particles, umbrella sampling, etc…)



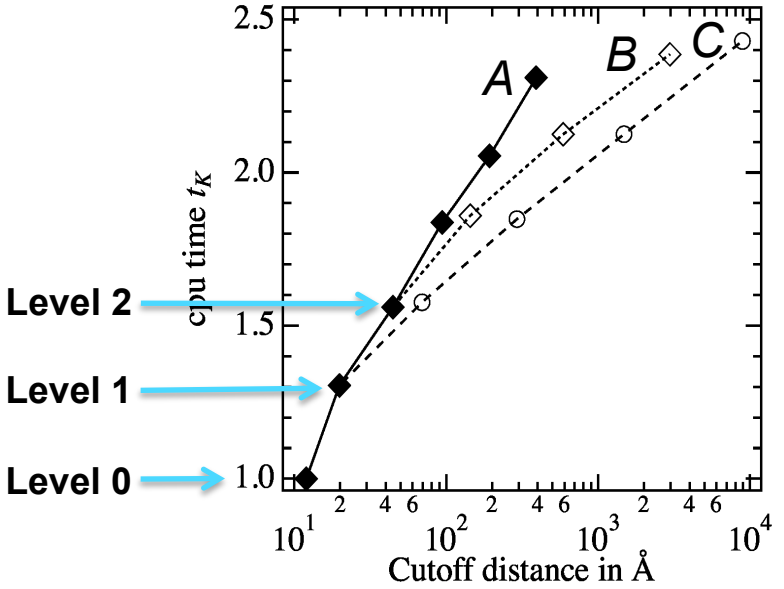
Masella *et al*, J. Comput Chem, 2013

MULTI-SCALE APPROACH EFFICIENCY

*M. Masella et al, J. Comput Chem, **32** (2011) 2664; ibid, **34** (2013) 1112*



Particle size	8	64
Box	8 V_0	64 V_0



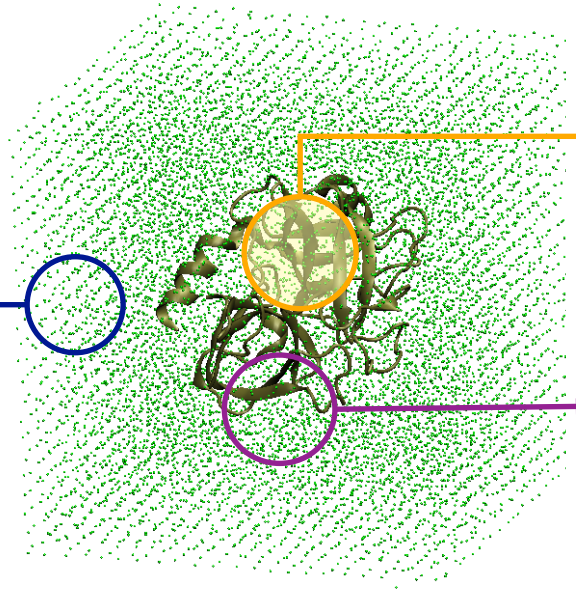
Boxes	0→1	1→2	2→3	3→4
A	x2	x2	x2	x2
B	x2	x2	x3	x4
C	x2	x3	x4	x5

Now, the bottleneck is handling the interactions within the solute…

THE CODE POLARIS(MD)

The code POLARIS(MD) © CEA/DSV

Masella, Mol. Phys., 2006 ; Masella *et al*, J. Comput Chem, 2008; *ibid*, 2011; *ibid*, 2013



Polarizable force-field

Energetic of creating a cavity within liquid water

⇒ **Hydrophobic effects**

Polarizable Pseudo Particles

⇒**Macroscopic laws of electrostatic**

Supported by the Exascale Computing Research Laboratory

energie atomique • energies alternatives

UNIVERSITÉ DE VERSAILLES SAINT-QUENTIN-EN-YVELINES

CODE EFFICIENCY ON MODERN SUPERCOMPUTERS

Reference : Intel Sandy-Bridge octo-cores 2.7 Ghz (CURIE)

All atom simulations+ PME summation techniques

A water box of 1000 molecules ➔ **3.0 ns per day** (1 cpu)
4.5 ns per day (4 cpu)

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

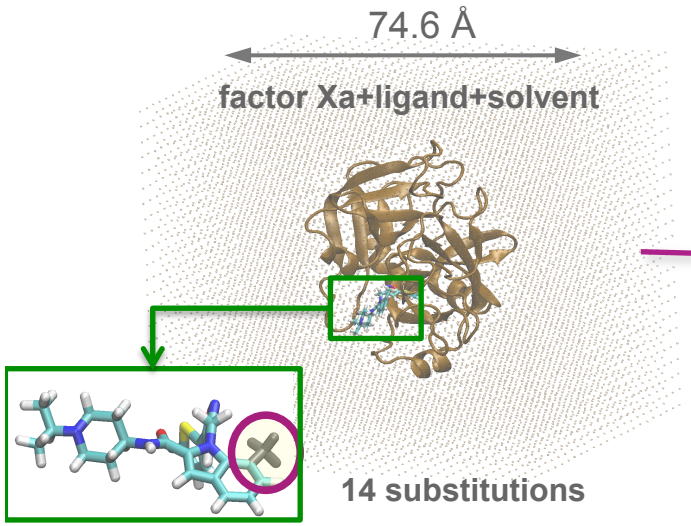
6000 particles + solute = 900 atoms ➔ **15 ns per day** (level 0 + 1 cpu)
13 ns per day (level 1 + 1 cpu)
11.5 ns per day (level 2 + 1 cpu)
9.5 ns per day (level 3 + 1 cpu)

Largest system simulated :

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

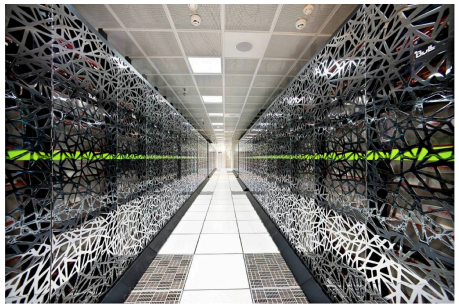
TOWARDS LARGE SCALE FREE ENERGY COMPUTATIONS

Free energy ΔΔG computations (TI schemes and 40 steps)

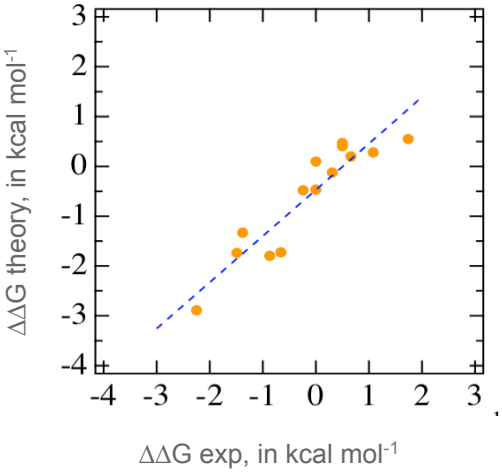


Solute = 3700 atoms
Solvent = 12400 particles (level 0)

TERA 100 (DAM/DIF)



560 simulations of 600 ps
2240 cores ≈ 32 h
(POLARIS(MD) 2010)
Today ➔6 h (Sandy Bridge)



POLARIS(MD) 2013 + solute *O(N²)* + 500 000 h CURIE :

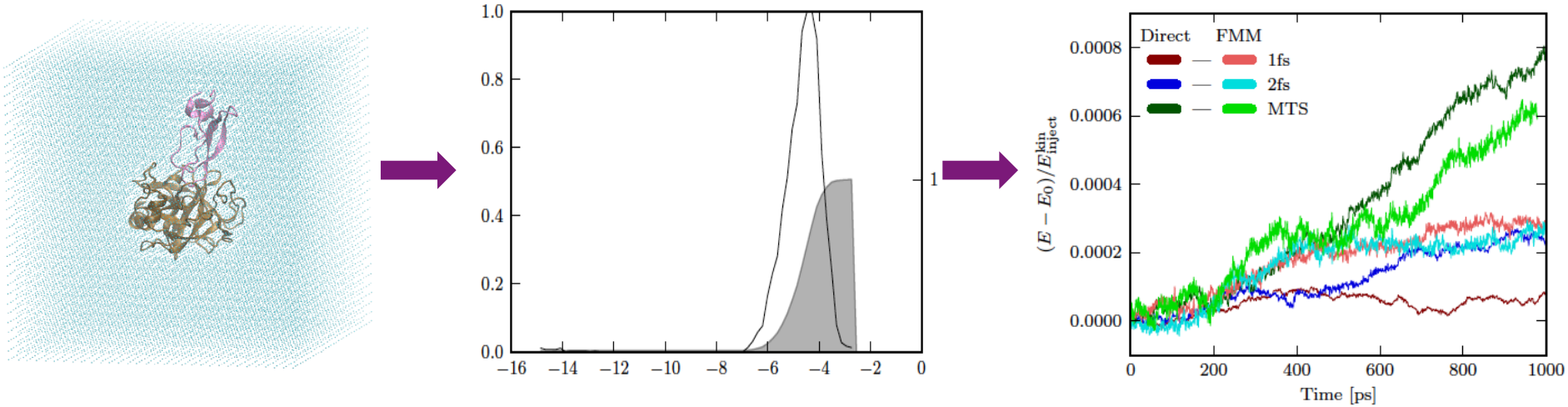
➔ 400 ΔΔG computations

A *O(N)* FAST MULTIPOLE METHOD

$$\phi(\boldsymbol{x}_b - \boldsymbol{x}_a) = \frac{q_a}{|\boldsymbol{x}_b - \boldsymbol{x}_a|} \quad \rightarrow \quad \phi(x_b - x_a) \approx \sum_{|\mathbf{n}|\leq p} \sum_{|\mathbf{m}|\leq p-|\mathbf{n}|} \frac{(-1)^{|\mathbf{n}|}}{\mathbf{n!m!}} r_b^{\mathbf{n}} r_a^{\mathbf{m}} \nabla^{\mathbf{n}+\mathbf{m}} \phi(z_B - z_A)$$

$$+ \quad \boldsymbol{\mu}_a = 2q_q^{\mu} \delta \boldsymbol{l}_a$$

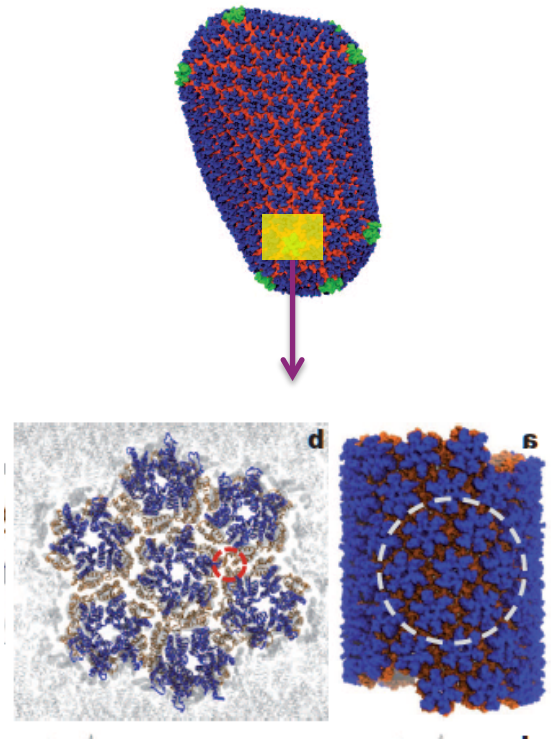
+ atoms organized via a kd-tree spatial decomposition

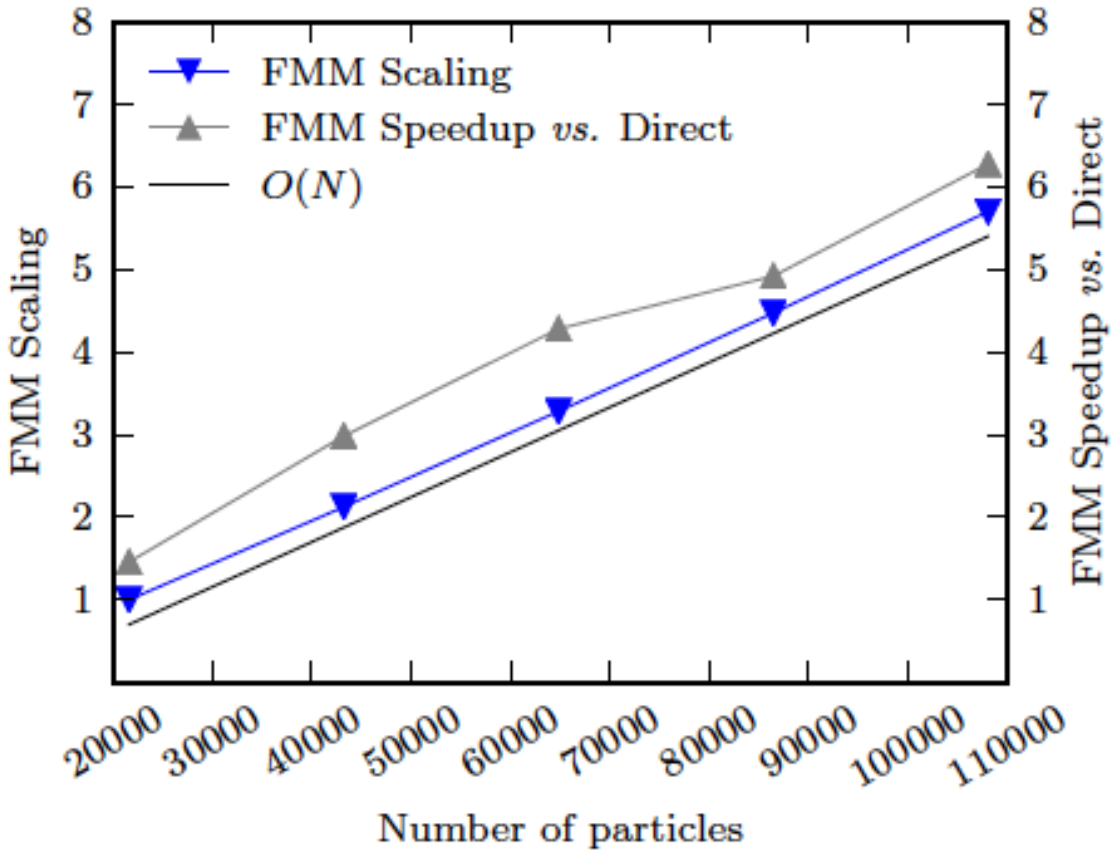


Coles and Masella, arXiv:1408.6527v1

FMM EFFICIENCY

The mature HIV-1 capsid

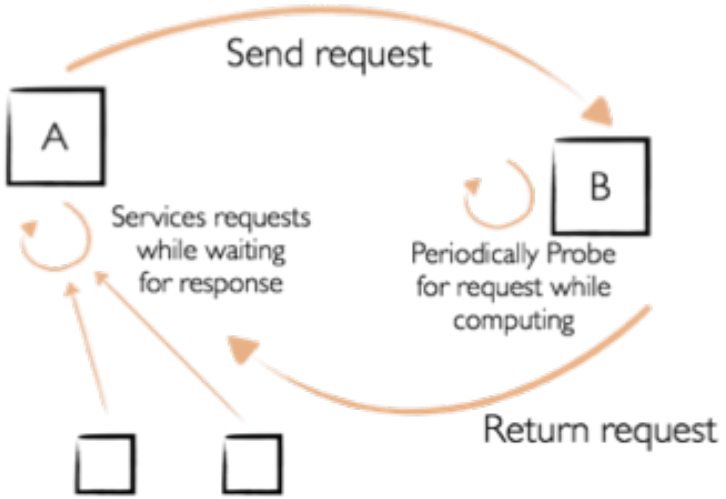




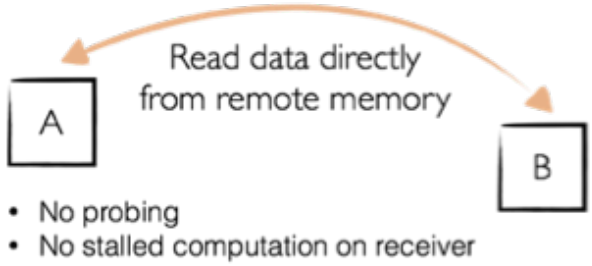
Coles and Masella, arXiv:1408.6527v1

TOWARDS A ONE SIDED COMMUNICATION SCHEME

Standard 2-sided communication scheme

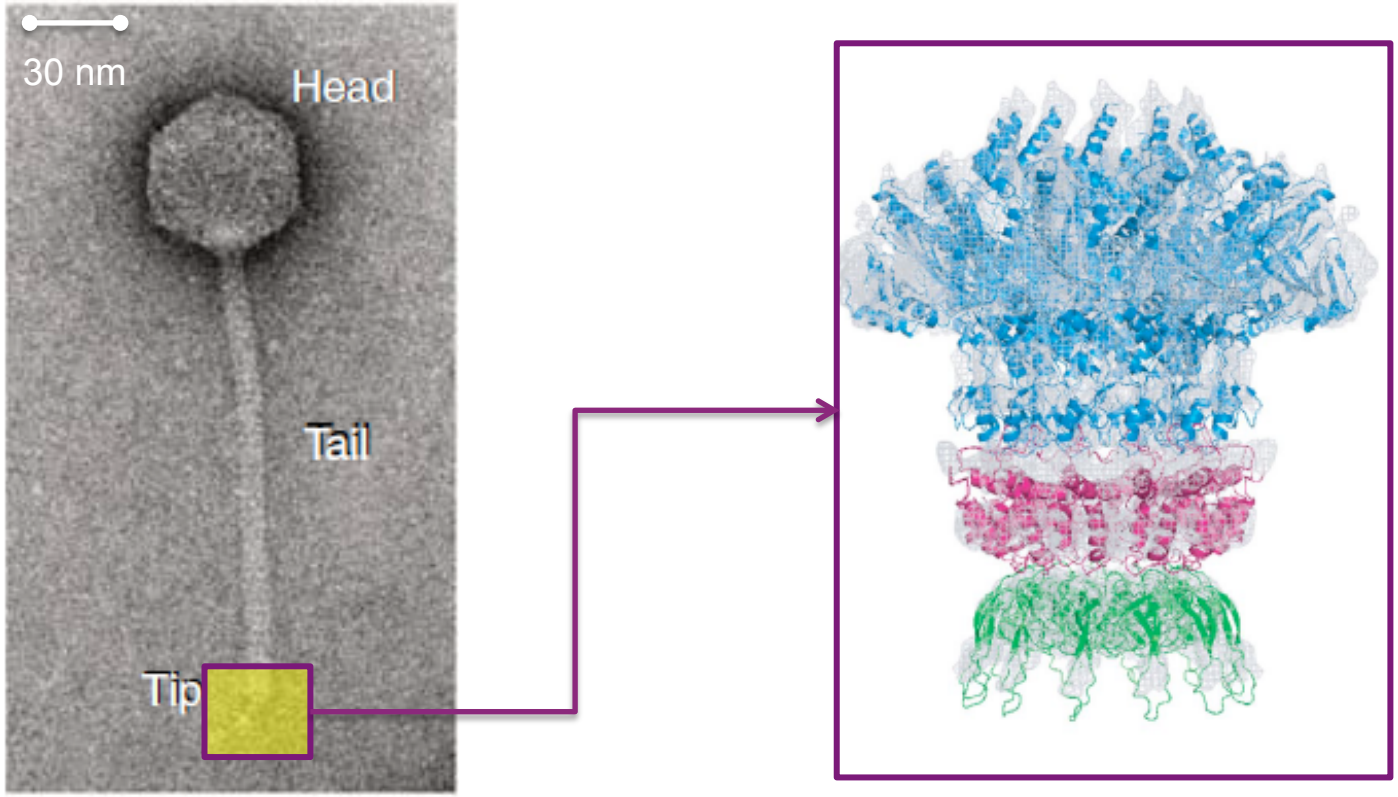


1-sided scheme, MPI v 3.0



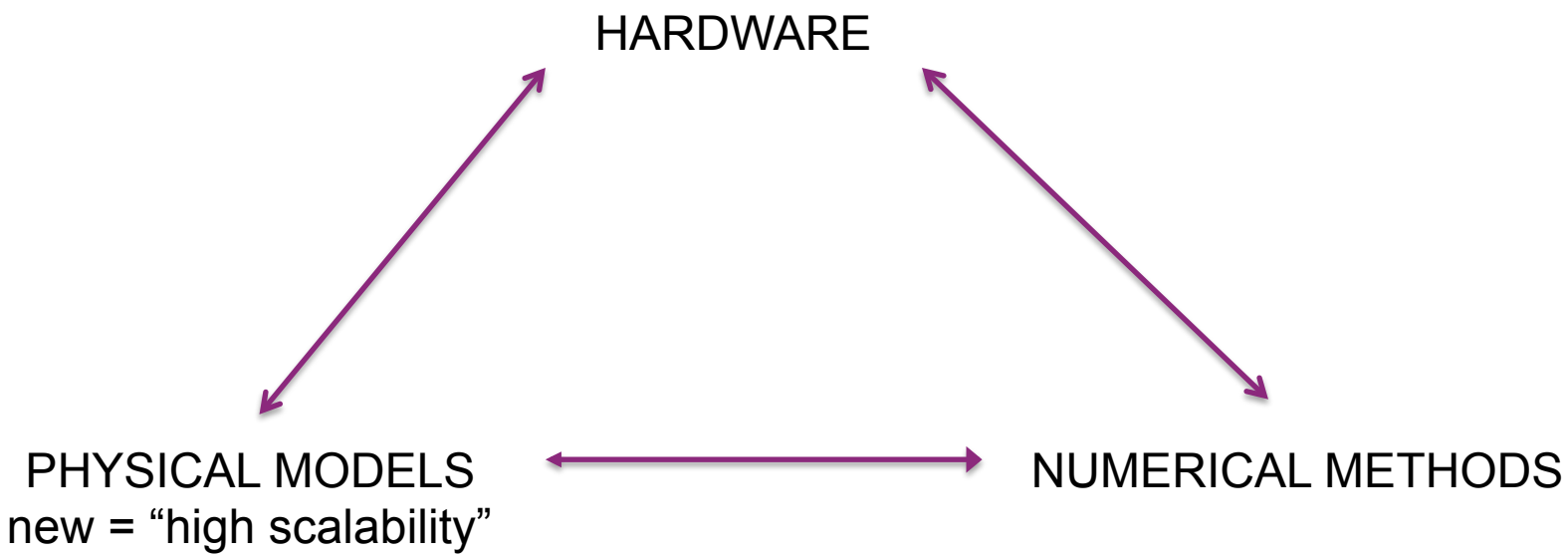
SIMULATING THE CELL MACHINERY

Bacillus subtilis bacteriophage SPP1 structure



Lhuillier et al, PNAS, ***106*** *(2008) 8507*

TOWARDS EXAFLOPIC SIMULATIONS



30 SEPTEMBRE 2014

Commissariat à l'énergie atomique et aux énergies alternatives
Centre de Saclay | 91191 Gif-sur-Yvette Cedex
T. +33 (0)1 69 08 99 62 | F. +33 (0)1 69 08 90 71

Life Science Division
IBITECS/SIMOPRO
Life chemistry laboratory

Etablissement public à caractère industriel et commercial | RCS Paris B 775 685 019