DE LA RECHERCHE À L'INDUSTRIE

cea

www.cea.fr

# SIMULER DES POLYMÈRES
# BIO ACTIFS

Forum TERATEC | Michel Masella | Basic Research Division

27 JUIN 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)

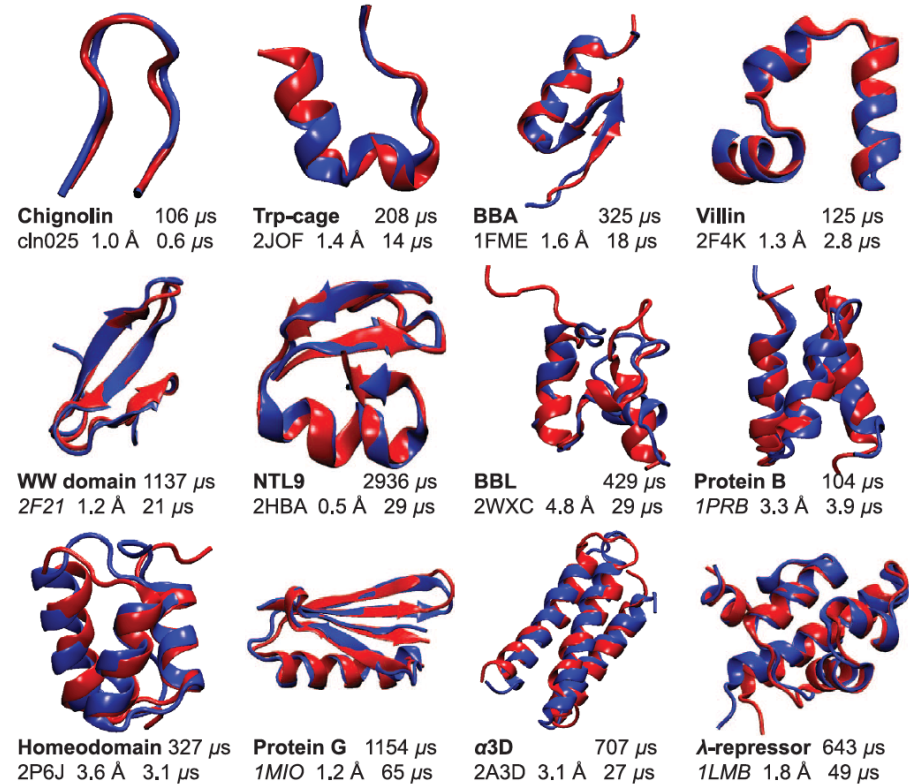**Example : simulating the mature HIV-1 capsid structure**
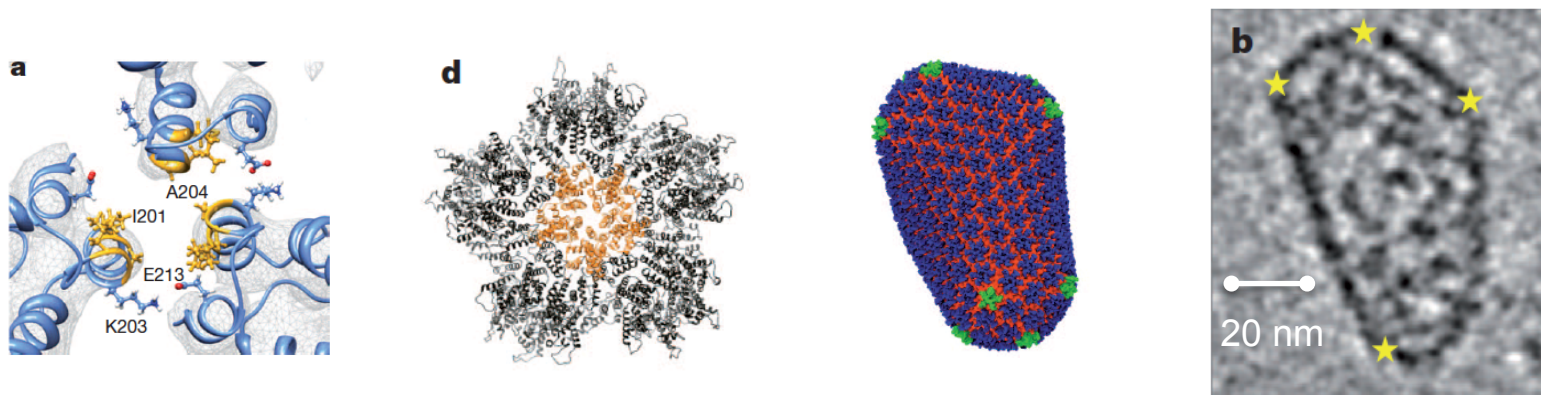


128 000 cores of the Cray "Blue Waters", 300 ns, about $60.10^6$ atoms ($50.10^6$ = water)

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

## Anton, a special-purpose machine for molecular dynamics simulation

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



| | | |
|---|---|---|
| **Chignolin** 106 μs | **Trp-cage** 208 μs | **BBA** 325 μs |
| cln025 1.0 Å 0.6 μs | 2JOF 1.4 Å 14 μs | 1FME 1.6 Å 18 μs |

**Villin** 125 μs
2F4K 1.3 Å 2.8 μs

**WW domain** 1137 μs   **NTL9** 2936 μs   **BBL** 429 μs   **Protein B** 104 μs
2F21 1.2 Å 21 μs   2HBA 0.5 Å 29 μs   2WXC 4.8 Å 29 μs   1PRB 3.3 Å 3.9 μs

**Homeodomain** 327 μs   **Protein G** 1154 μs   **α3D** 707 μs   **λ-repressor** 643 μs
2P6J 3.6 Å 3.1 μs   1MIO 1.2 Å 65 μs   2A3D 3.1 Å 27 μs   1LMB 1.8 Å 49 μs

## How fast folding proteins fold

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

## The main drawback of MM approach: accuracy

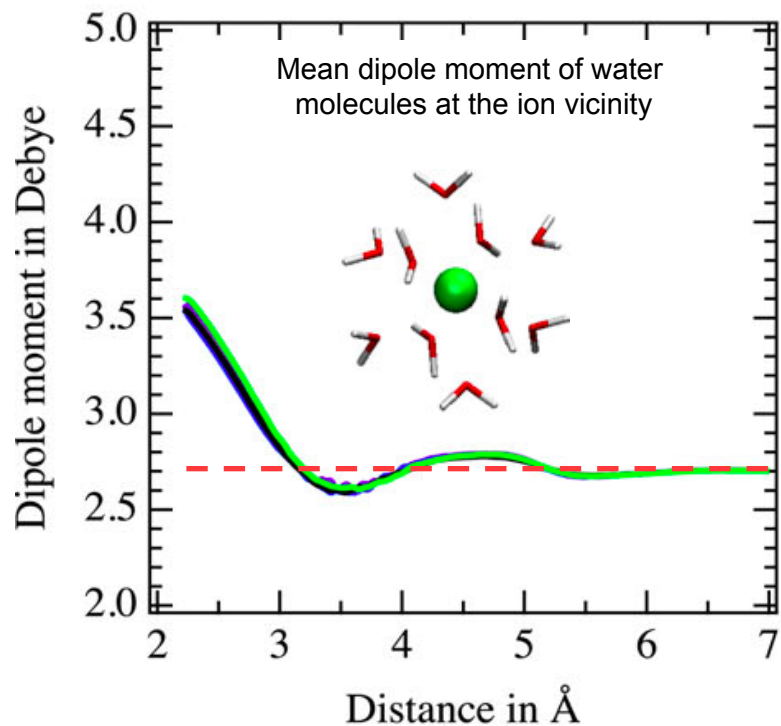Common force-fields are based on additive potentials (CHARMM, AMBER, OPLS …)

1 - The model parameters are constant along an MD trajectory

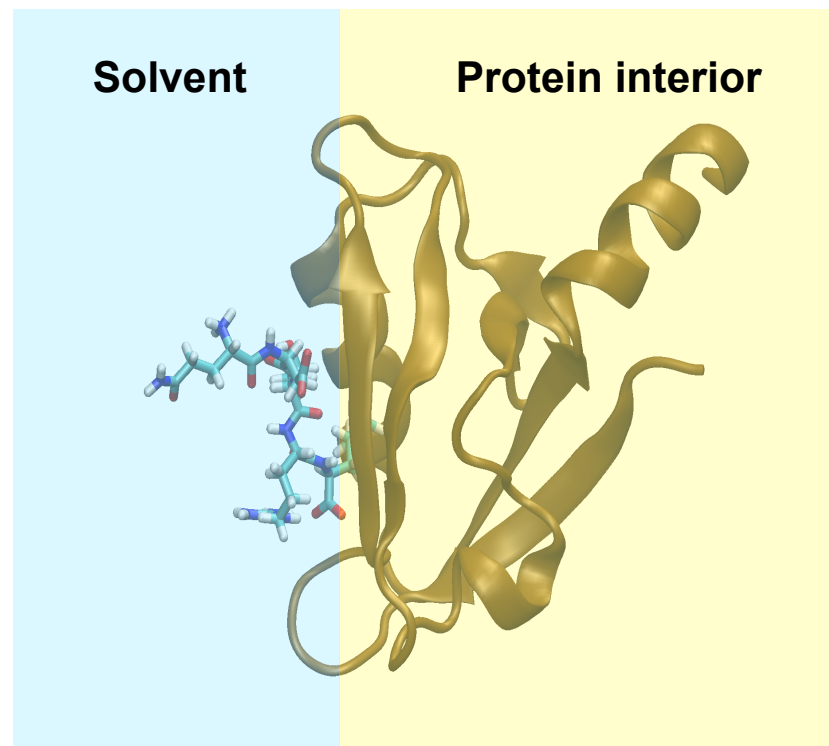2 – Reliability of such an approach when mimicking microscopic electrostatic interactions

➜  $U_{Coulombic} = \sum_{i<j} \dfrac{\delta_i \delta_j}{r_{ij}}$  ➜  $\delta_i$ are static charges
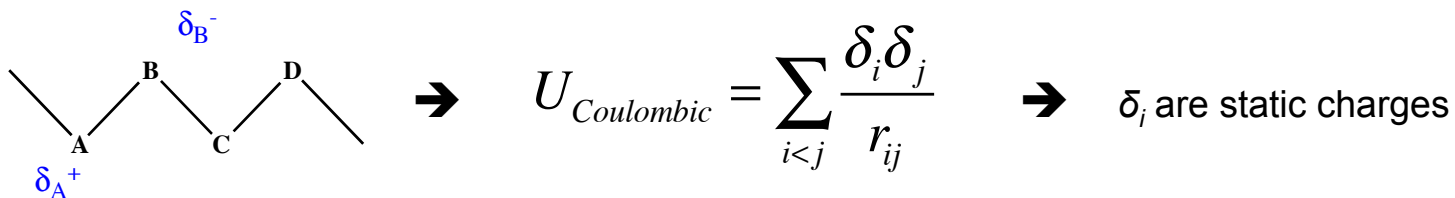
3 – **Experimental results** are mainly used for assigning/refining parameters, transferability ?

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

## However, what about interfaces ?

**The case of Cm(III) interacting with water**

**All protein/ligand interactions correspond to an interface problem**



Mean dipole moment of water molecules at the ion vicinity

Dipole moment in Debye

Distance in Å

Solvent

Protein interior

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

**The three next major milestones in molecular modeling**

- We need more accurate models to quantify ligand/target binding, in particular to address interface problems (i.e. we need more physics)

- How to assign parameters when facing poorly documented new molecules ?

- We need also high throughput and low cost modeling approaches, using standard computational units and providing significant results at the week scale

**Addressing these points will lead not only to perzonalyzed medecine applications, but also to address toxicology, pollution and all related problems, in particular to assess the properties of non synthetized (non exisiting) new families of molecules.**

# Improving the functional form of molecular modeling Hamiltonians

**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**

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

$\mathbf{T}$ is the dipolar tensor and $\alpha_i$ is the center $i$ polarizability, usually isotropic
$\mathbf{E}_i$ 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**

2nd coordination sphere

1st coordination sphere
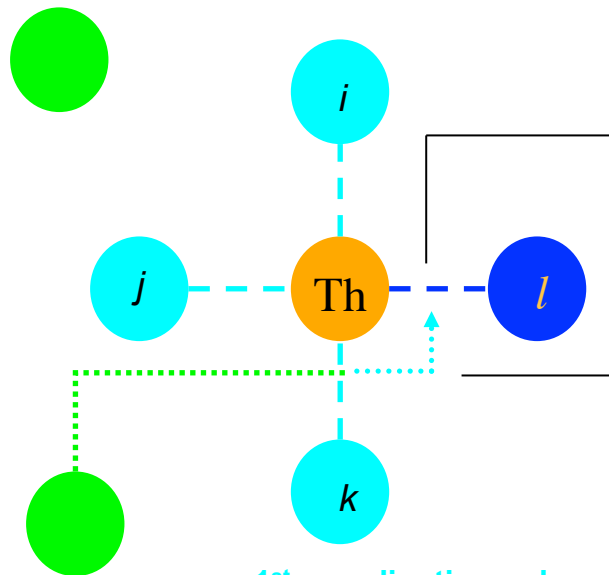
$$U_{ct} = \sum_{l \in water} D_e^c \exp(-\beta \cdot r_{Th-l})$$

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

**Seems computationally demanding, however scales as _O(N)_**

# An quantum ab initio based strategy to assign force field parameters

**The most precise theoretical methods in chemistry are the quantum (QM) ones based on solving the Schrödinger equation :**

$$H\Psi = E\Psi$$

The most precise methods allowing to reliably account for electronic correlation scale as $O(N^{p\geq2})$, N being the number of electrons (62 electrons for a system of 14 atoms) ➜ **heavy computations**

Today, it is possible to perform hundreds of high end quantum computations at the week scale :

1 - It is thus possible to generate large enough target data sets to assign force field parameters;

2 - **the typical precision on the QM energies is << 1 kcal mol$^{-1}$.**

➜ **It is now possible to precisely investigate the microscopic interactions of any kind of (small enough) molecules, regardless they are synthetized or not…**

Benchmark binding energies of ammonium and alkyl-ammonium ions interacting with water. Are ammonium-water hydrogen bonds strong?



Environmental effects

weakend the NH-O hydrogen bonds
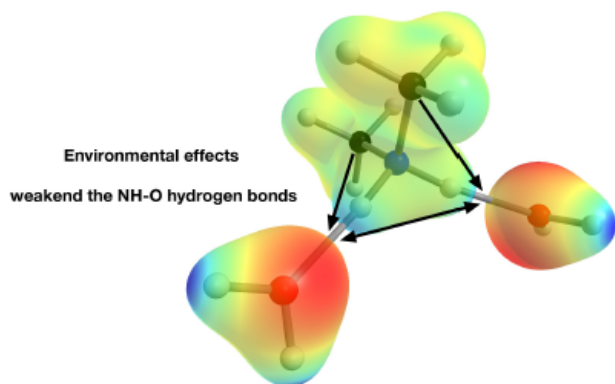
**V. Vallet and M. Masella, CPL, 618 (2015) 168**

$$\Delta H_{solv}(\mathrm{H}^+) = \Delta H_{g \to aq}(\mathrm{PH}) - \Delta H_{g \to aq}(\mathrm{P}) + \Delta H_g^{prot}(\mathrm{P}) + \Delta H_{aq,\mathrm{dissociation}}^0(\mathrm{PH})$$
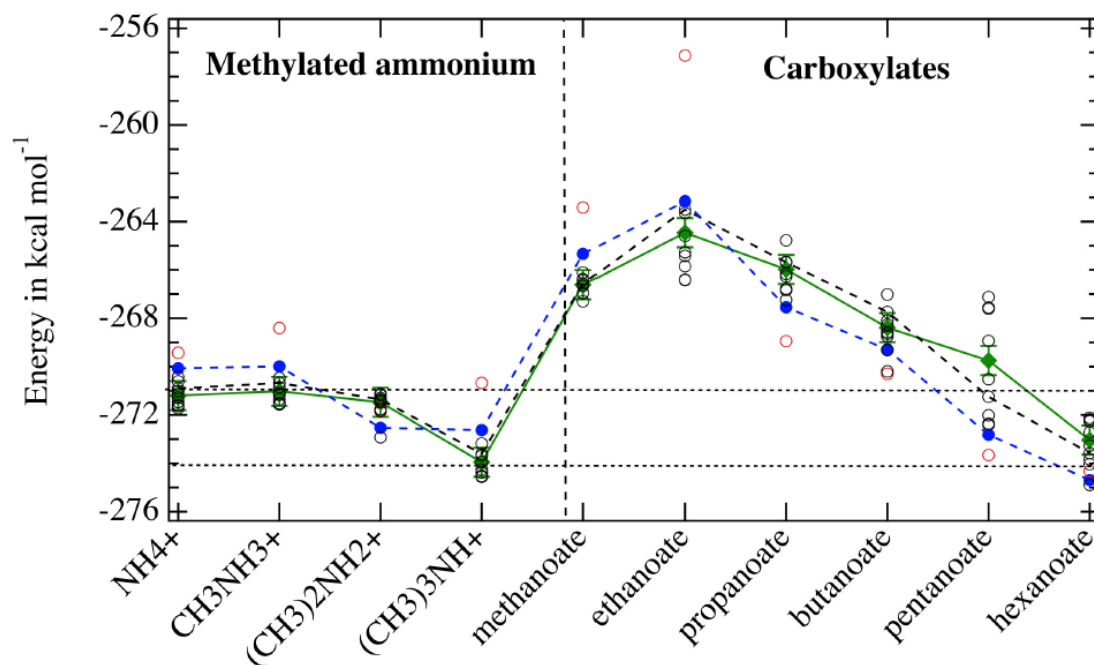
# Towards a high througput (and low cost) molecular modeling approach :
# A multiscale solvent coarse grained model

**Explicit solvent approach, need of FFT !**
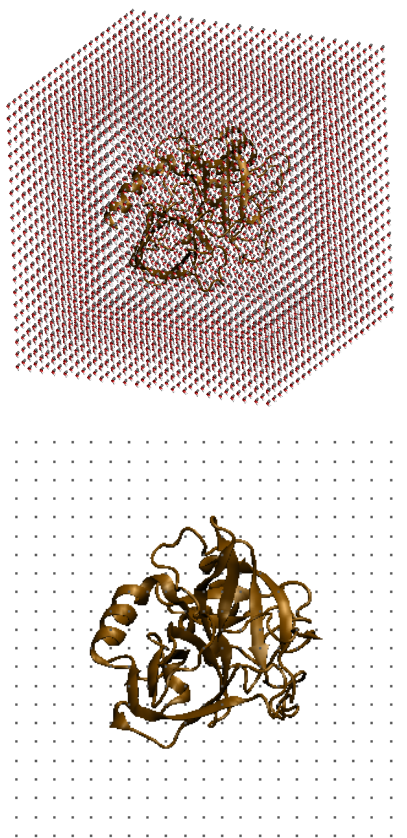
**Coarse grained approach**
Efficiency and microscopic solvation properties
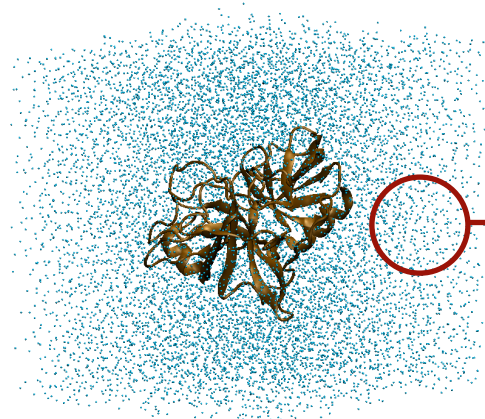


$$\alpha_S = \frac{1}{4\pi}\left(1 - \frac{1}{\varepsilon_s}\right) \times \frac{1}{\rho_s}$$

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

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

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

**Implicit solvent approaches**
**Poisson-Boltzman**
**Warshell's grid approach**

*Haduong et al, J. Chem. Phys., **117** (2002) 541*
*Masella et al, J. Comput. Chem., **29** (2008) 1707*

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

➔ **a multi-scale coarse grained approach to handle long tail bulk electrostatic**
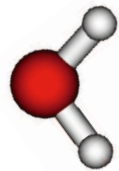
An explicit water molecule
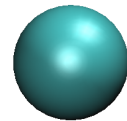
A polarizable pseudo-particule
= a water molecule

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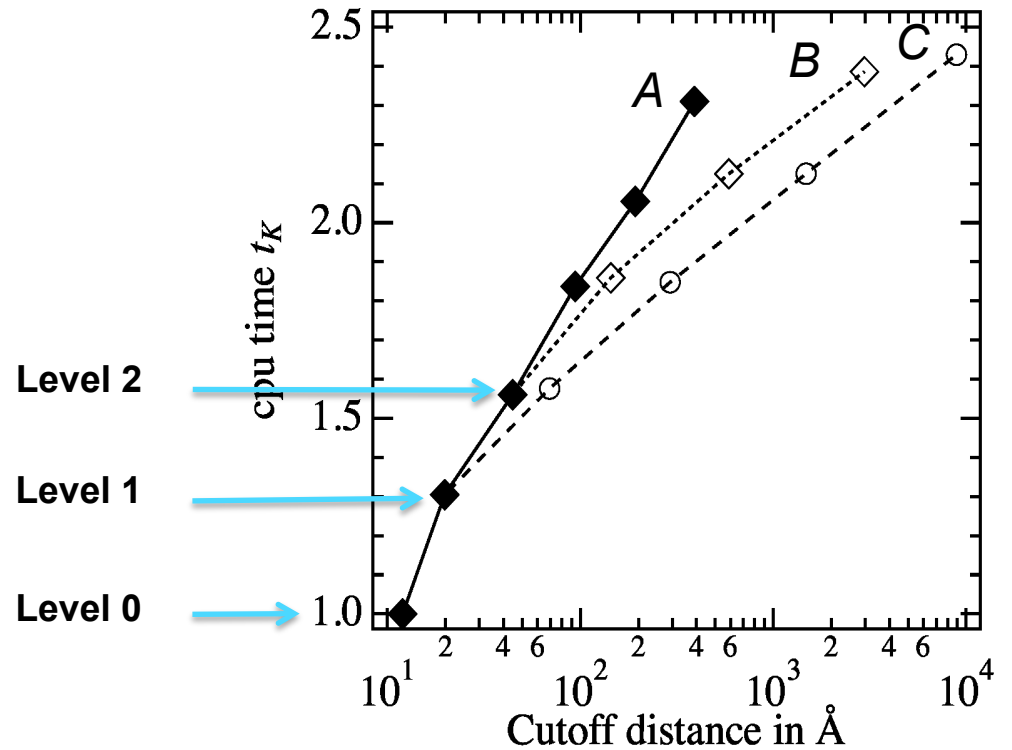= a volume element of liquid water

*Pseudo-particule
= a water molecule*

*Pseudo-particule
= volume element made
of 8 water molecules*

*Pseudo-particule
= volume element made
of 64 water molecules*

*etc*

**Solute**

**Level 0**

**Level 1**

**Level 2**

**Level *n***

$R \approx 12\ \text{Å}$

$R \approx 30\text{-}40\ \text{Å}$
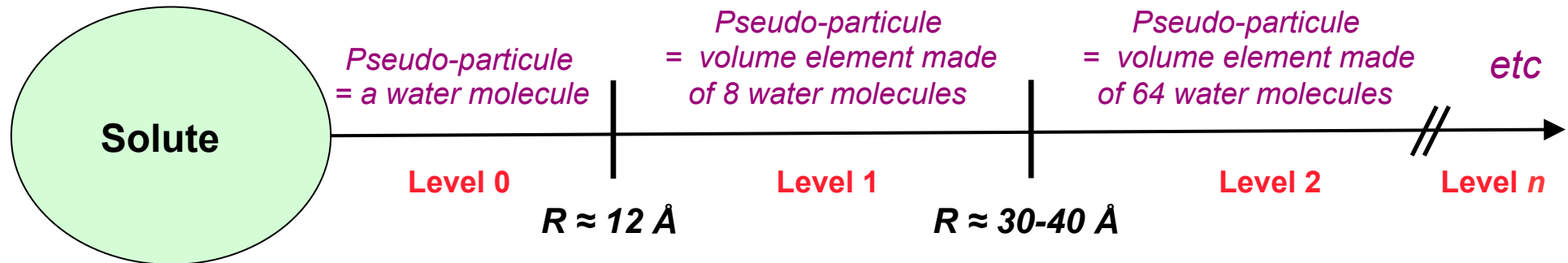
*M. Masella et al, J. Comput Chem, **32** (2011) 2664; ibid, **34** (2013) 1112*
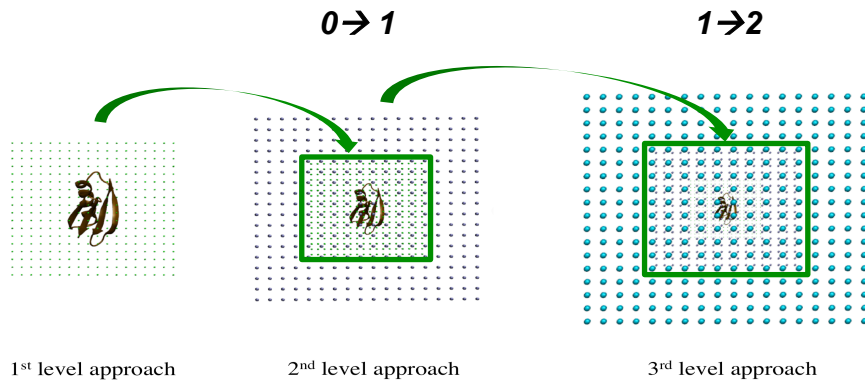


| Boxes | 0→1 | 1→2 | 2→3 | 3→4 |
|-------|-----|-----|-----|-----|
| A | x2 | x2 | x2 | x2 |
| B | x2 | x2 | x3 | x4 |
| C | x2 | x3 | x4 | x5 |

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

**Potential of mean force of Cl⁻/Cl⁻ and the coarse grained approach level**

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



$$\frac{q_1 q_2}{4\pi\varepsilon_s R_{12}}$$
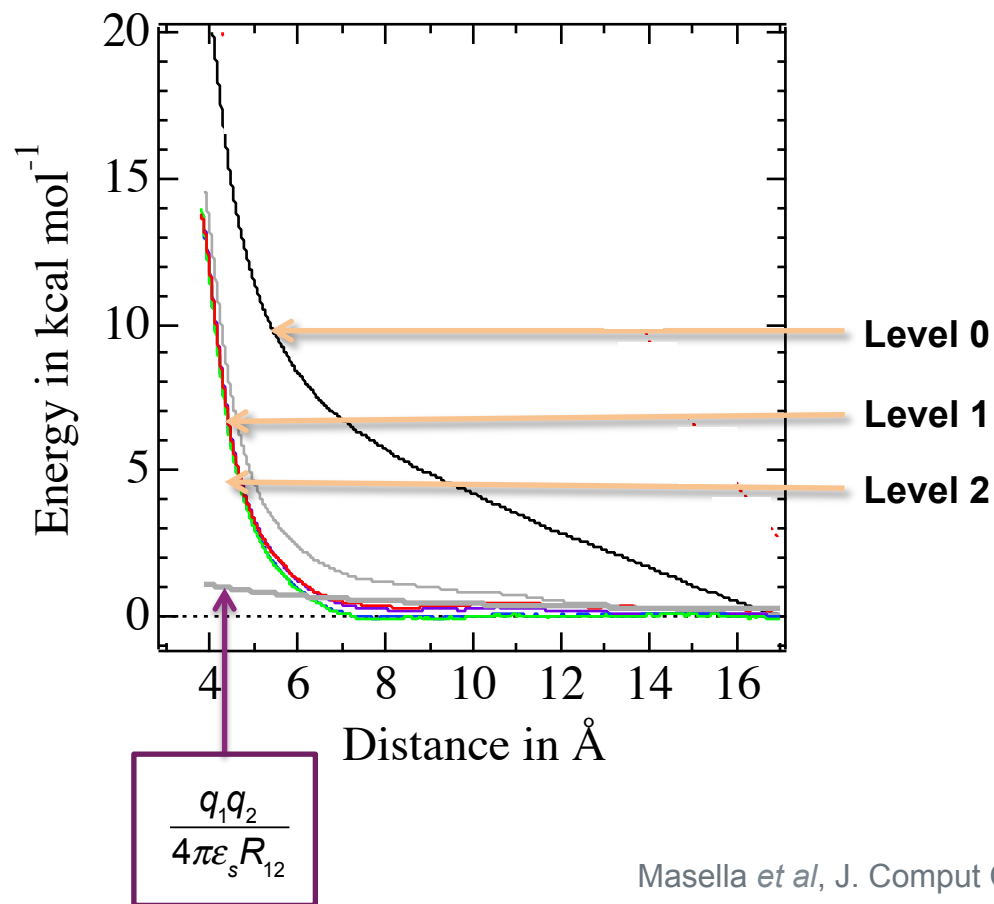
Masella *et al*, J. Comput Chem, 2013

**Free energy cost ∆G for creating a cavity within liquid water**

D. Chandler, Nature, **437** (2005) 640

*« Dry »*

*Enthalpy driven*

$\gamma$

*Entropy driven*

*« Wet »*
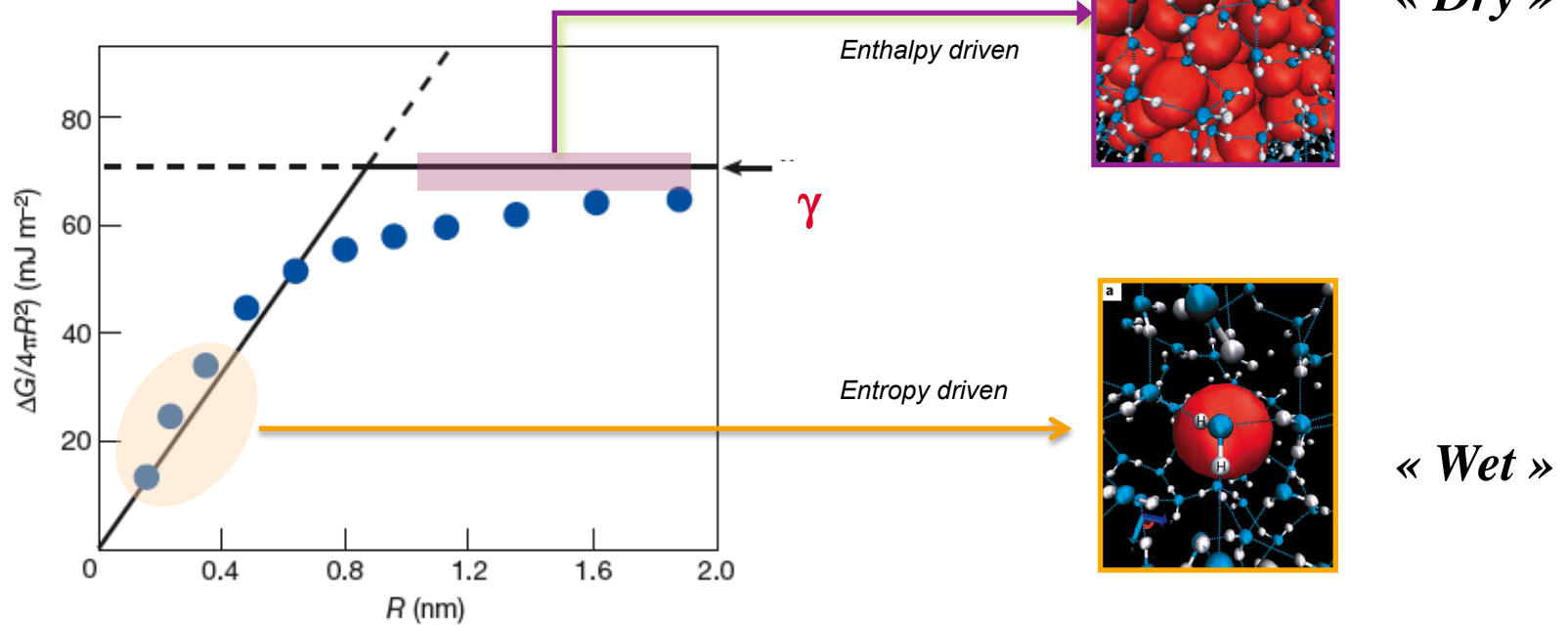
Taken into account in the model using

$$U_{pp}^{density} = \sum_{1 \le i \le N_s} \varepsilon_s^0 \left( n_i^0 - \overline{n}^0 \right)^2 + \varepsilon_s^1 \left( n_i^1 - \overline{n}^1 \right)^2$$

# An *O(N)* multi-scale *N*-body approach for simulating polarizable microscopic systems

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** (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 !!!**

$$\phi(\boldsymbol{x}_b - \boldsymbol{x}_a) = \frac{q_a}{|\boldsymbol{x}_b - \boldsymbol{x}_a|}$$

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

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

+ atoms organized via a kd-tree spatial decomposition

Mean error on the forces

Calcul électrostatique avec la FMM

Temps normalisé — No. Atomes

Cube d'eau
Régression linéaire
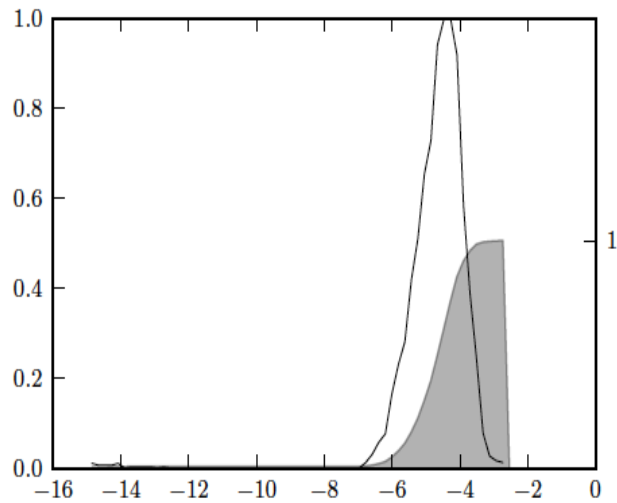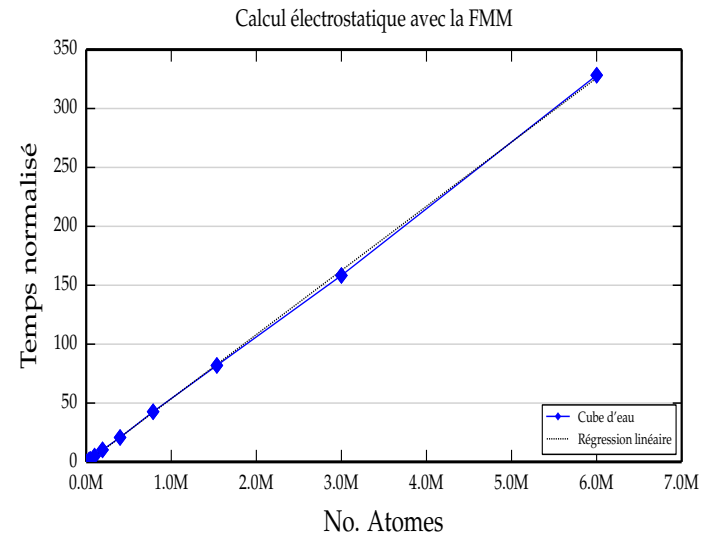
Strong Scaling of Electrostatic Energy for PMV Capsid
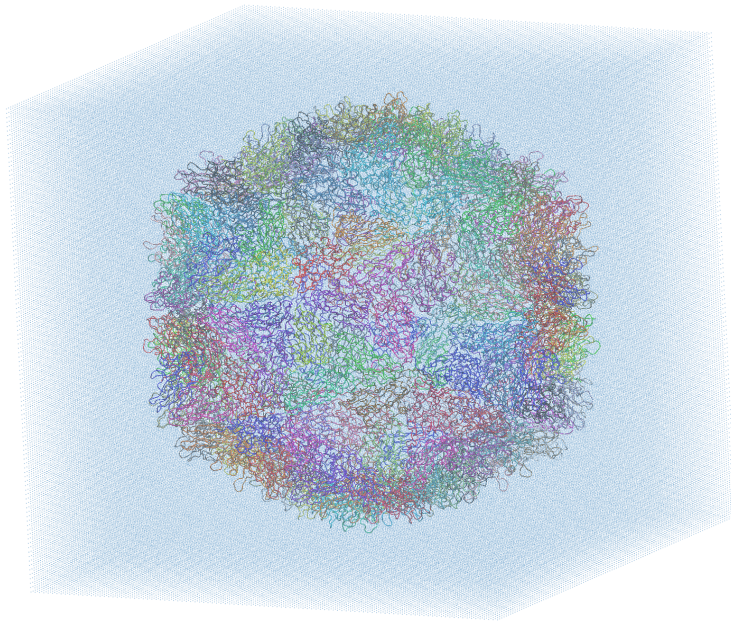
**Mosaic Panicum virus capsid**
0.6 M atoms + 1.8 M coarse grained
solvent particules
Equivalent to a 6.5 M atom systems
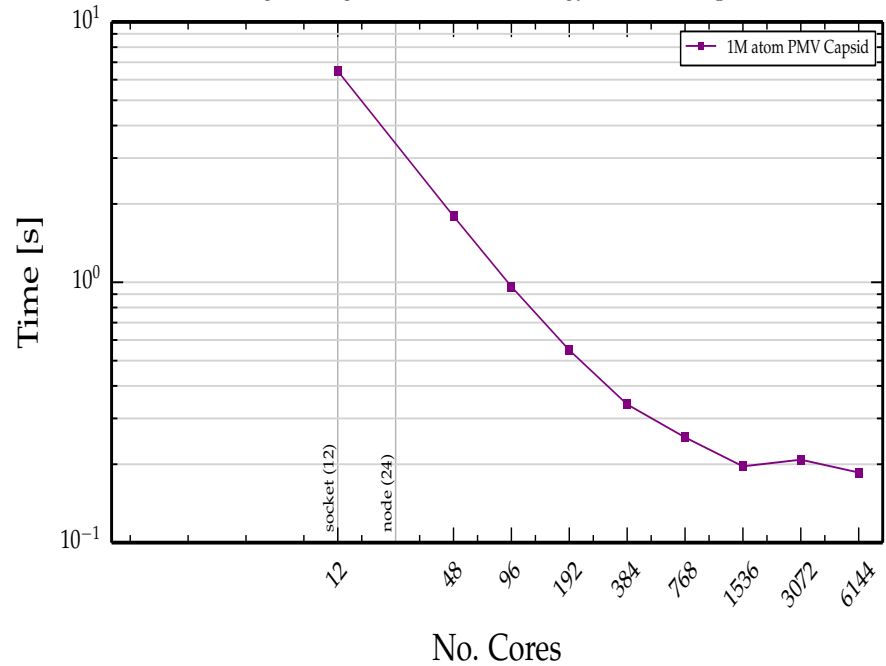
**Method scalabilty**
Hazelhen supercomputing system
(HLRS, Stuttgard, Gemany)

## 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**
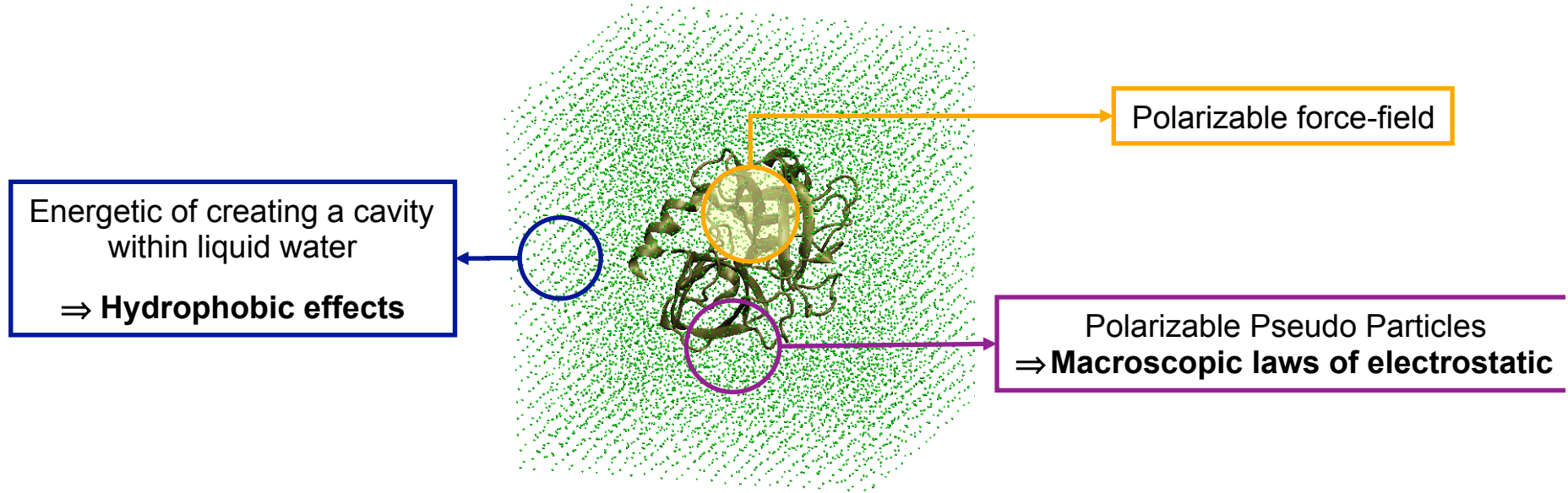
**Supported by the Exascale Computing Research Laboratory**

(intel)

cea

energie atomique • energies alternatives

UNIVERSITÉ DE VERSAILLES
SAINT-QUENTIN-EN-YVELINES

# Synthetic urea-based polymers and as new cancer breast bullets

**Peptides and proteins**
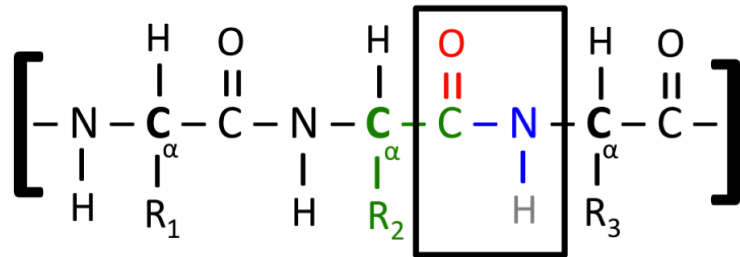
**Synthetic urea-based polymers**
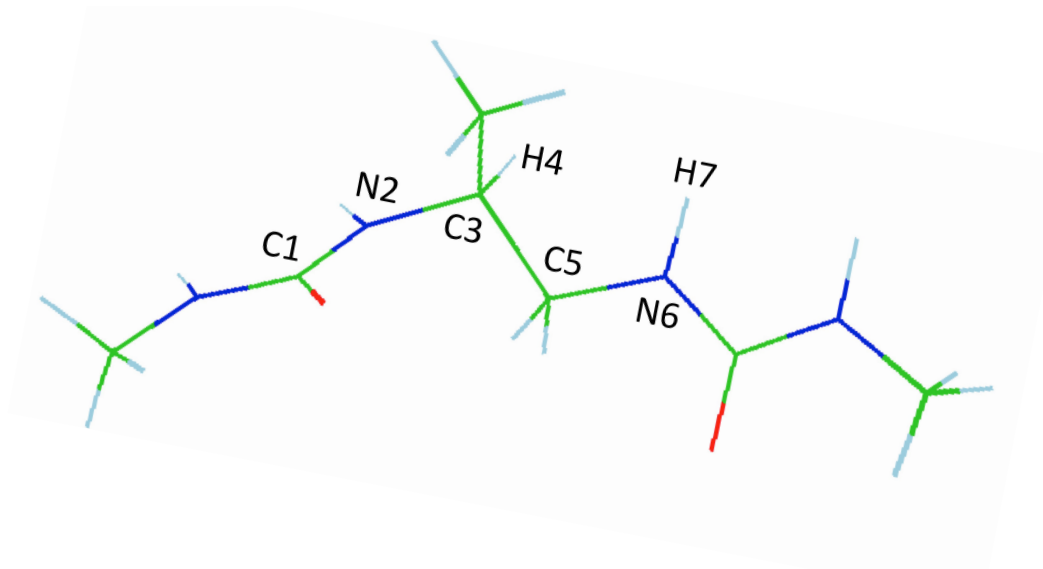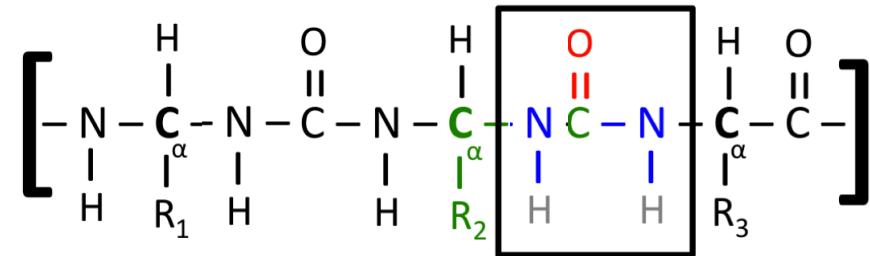
Guichard's team, University of Bordeaux
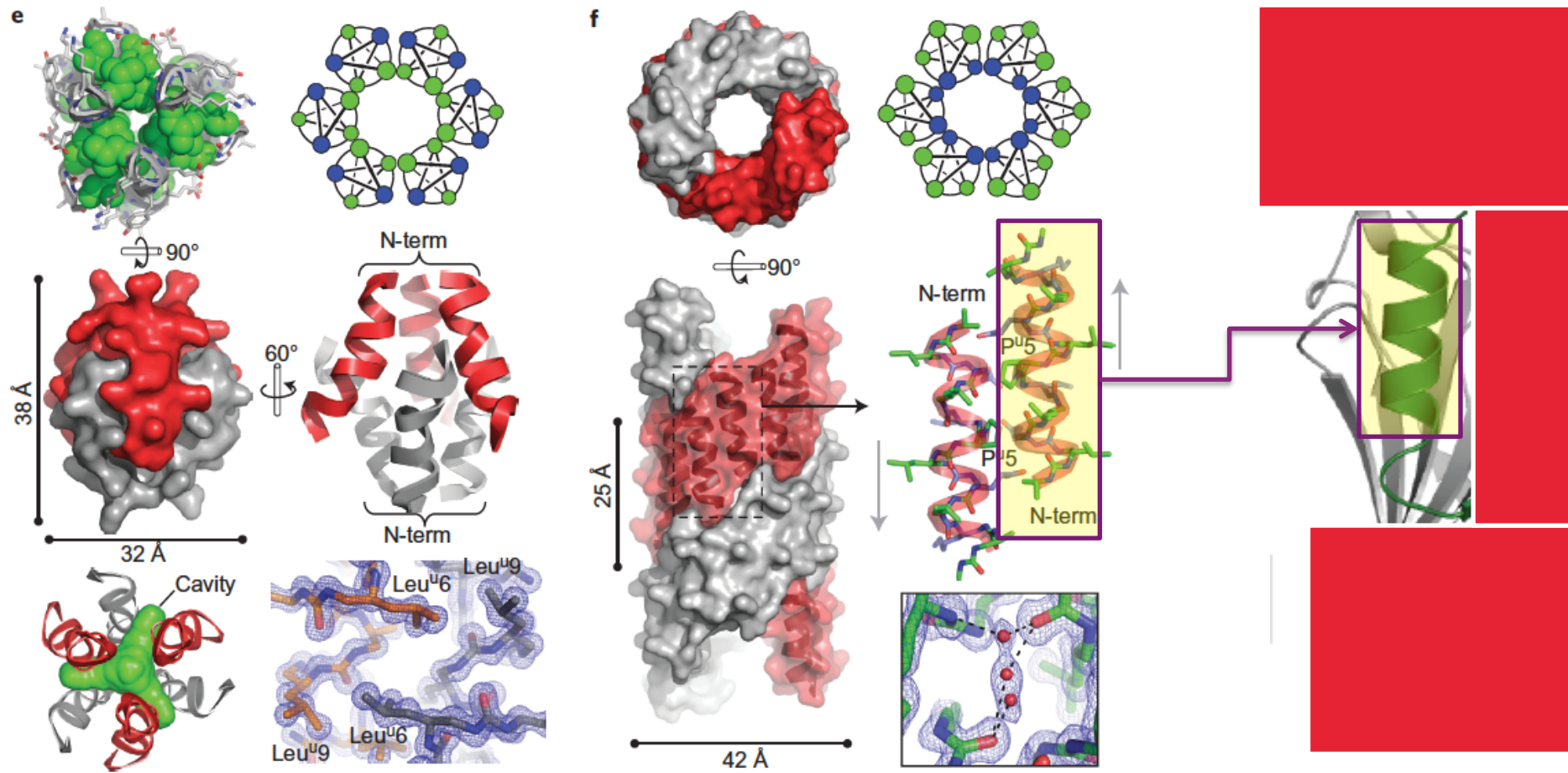
*Collie et al, Nature Chemistry, 15 (2015) 871*

## Xray

(1.9 Å)

**3nM**
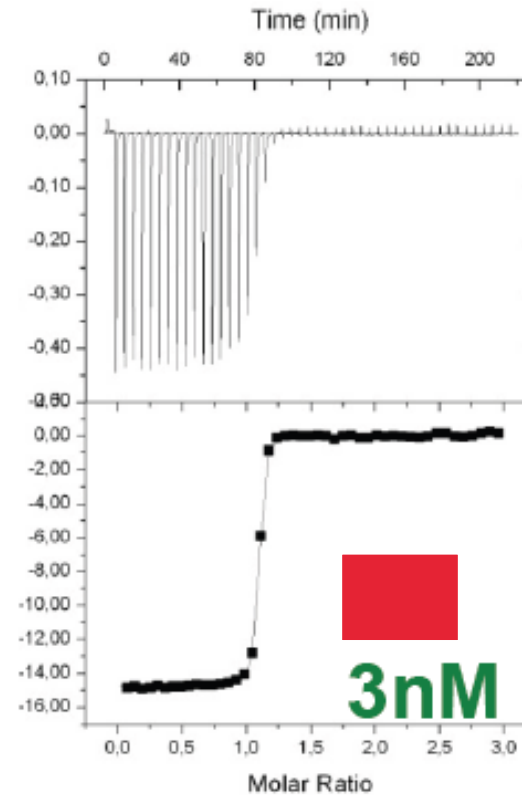
Peptide ⬛ ligand developed by the ⬛ teams

# Simulating synthetic bio mimetic polymers
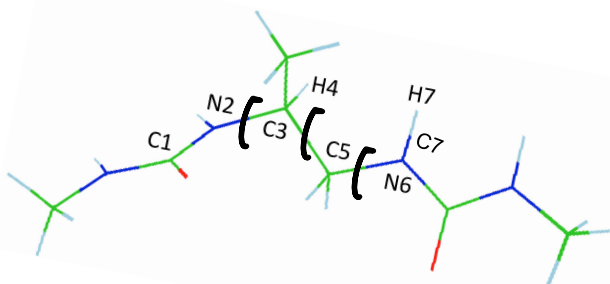
Map (1) : dihedral angle C1-N2-C3-C5/N2-C3-C5-N6
Map (2) : dihedral angle N2-C3-C5-N6/C3-C5-N6-C7
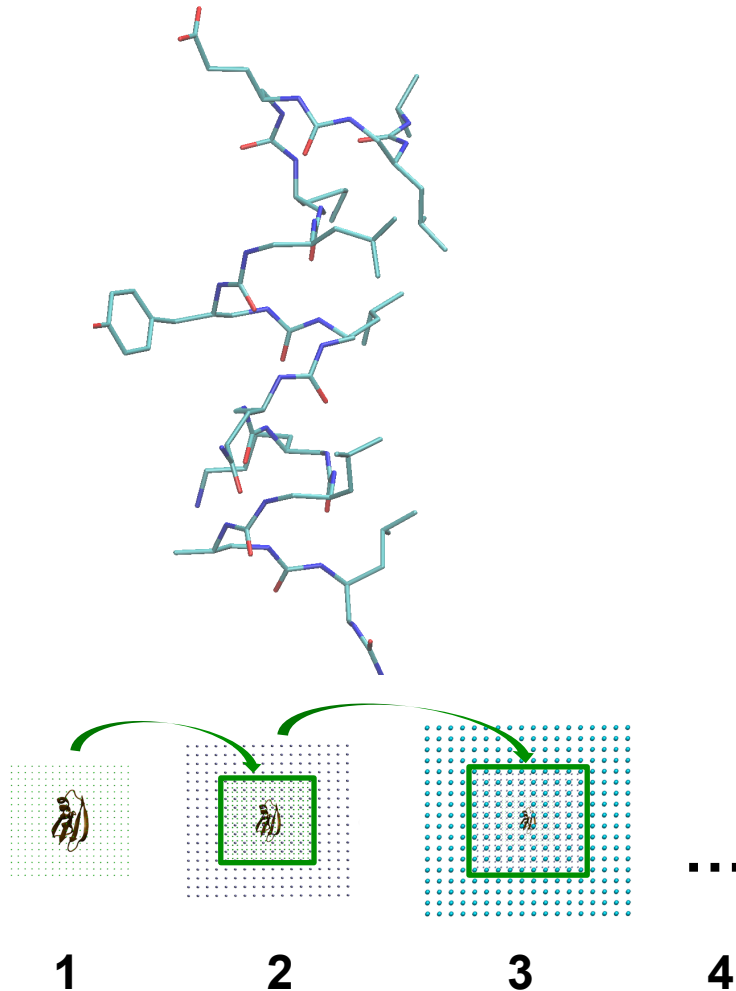Planes N-CO-N in their helix orientation

**Dihedral map (1)**

**Dihedral map (2)**
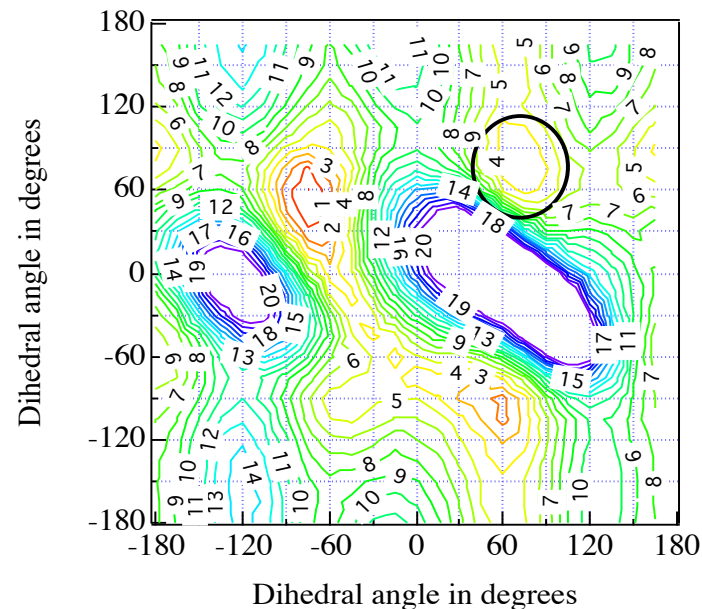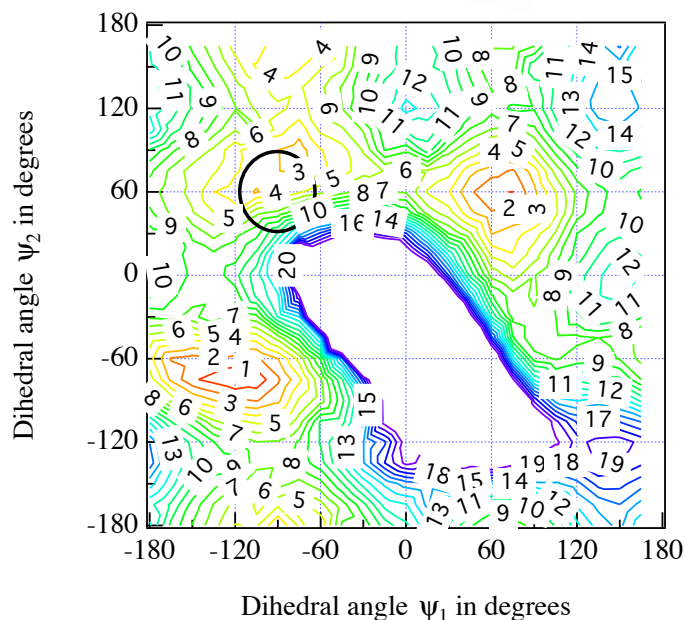
→ **Two Φ/ψ-like maps for urea based polymers (precision < 1 kcal mol⁻¹)**

**For each solvent level N, 10 MD simulations of the urea helix**

**The urea helix is embedded in solvent boxes, levels N = 1 to 4.**

**P : percentage of surviving helix ("Cα" rmsd < 1.0 Å at the end of 25 ns simulations).**



**Helix stable in pure aqueous phase (80%)**

**Stability at interfaces ?**
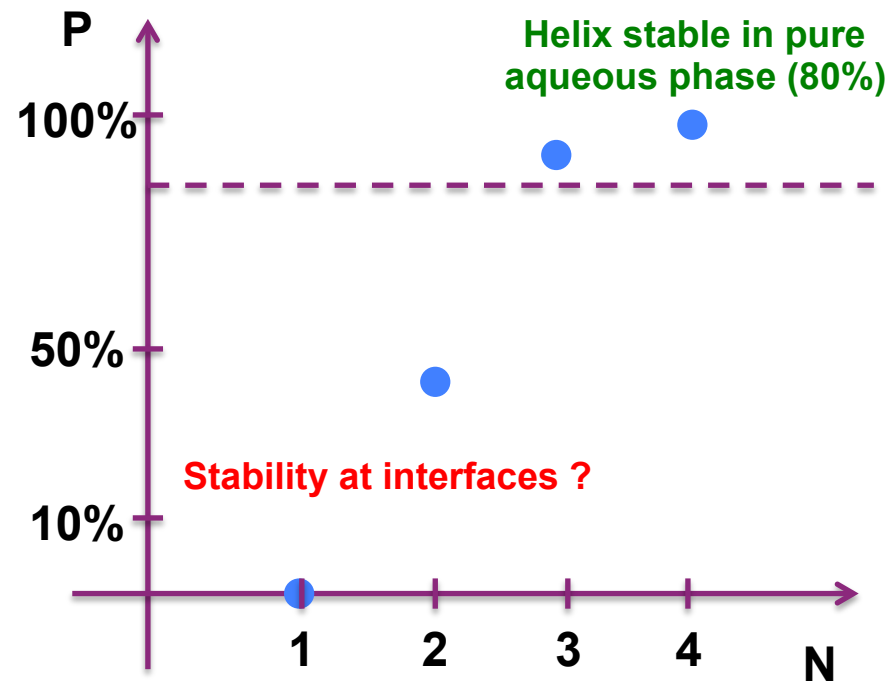
**RMSD temporal evolution vs solvent box level**
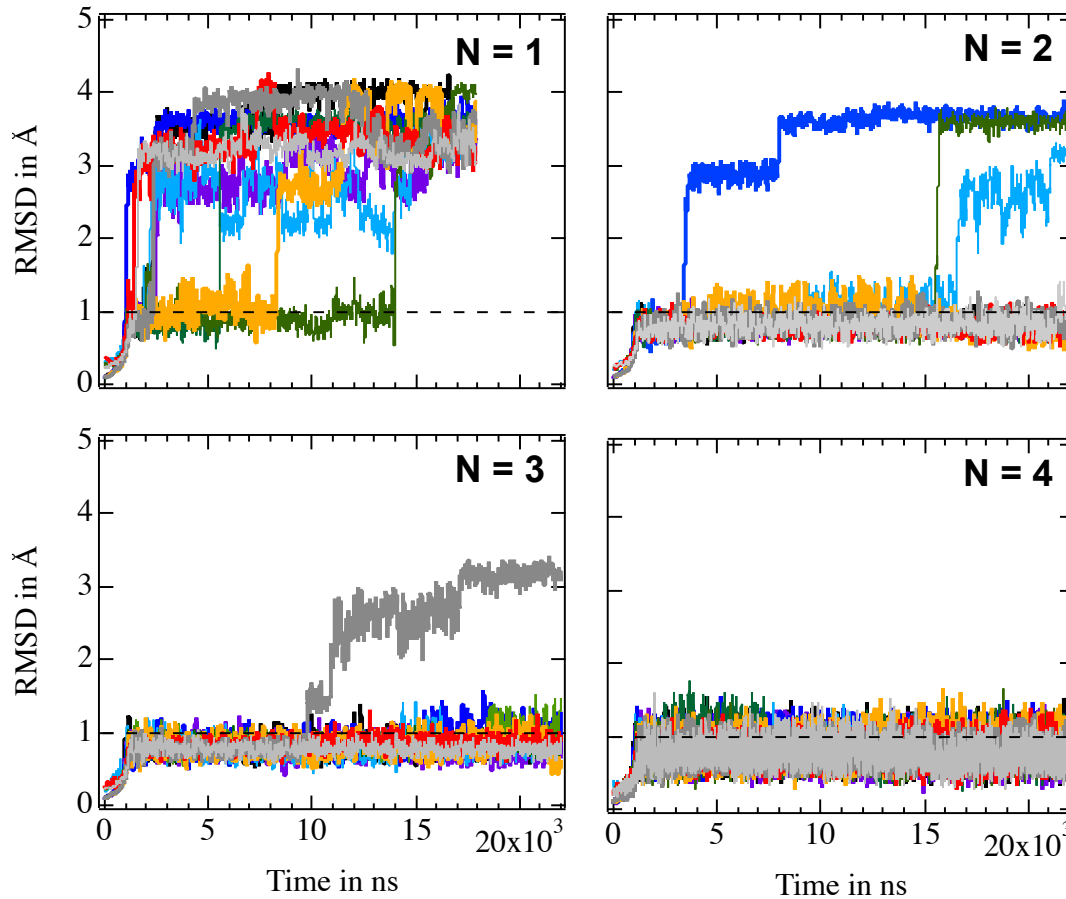


**RMSD vs dipole**
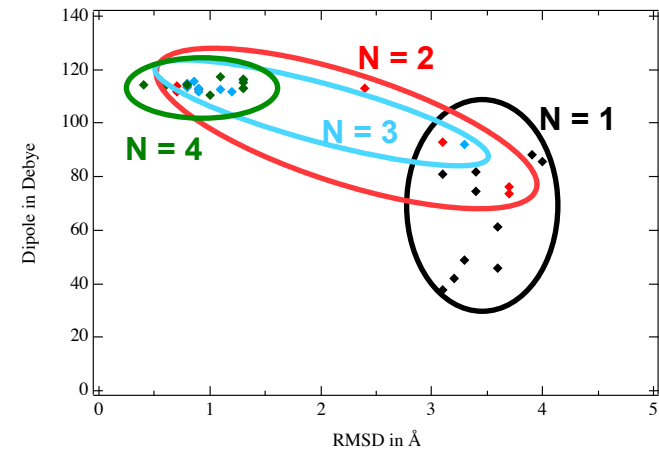


$$G_{Onsager} = -\frac{1}{4\pi\varepsilon_0} \times \frac{\varepsilon-1}{2\varepsilon+1} \times \frac{M^2}{R^3}$$
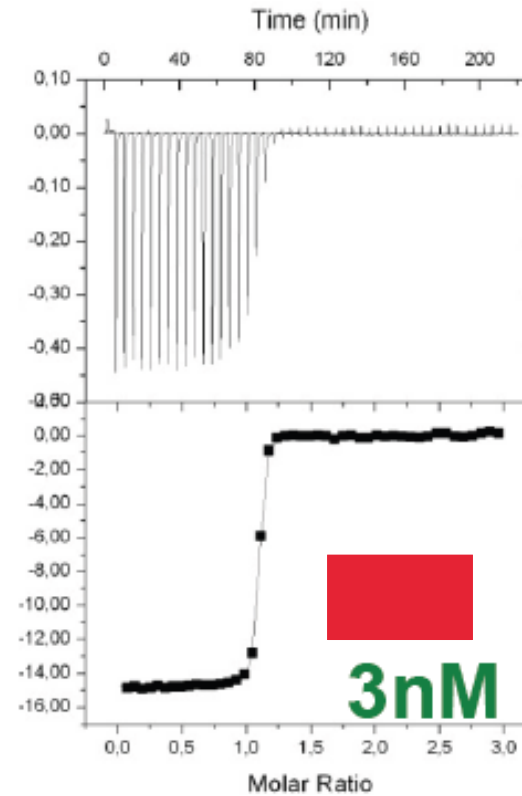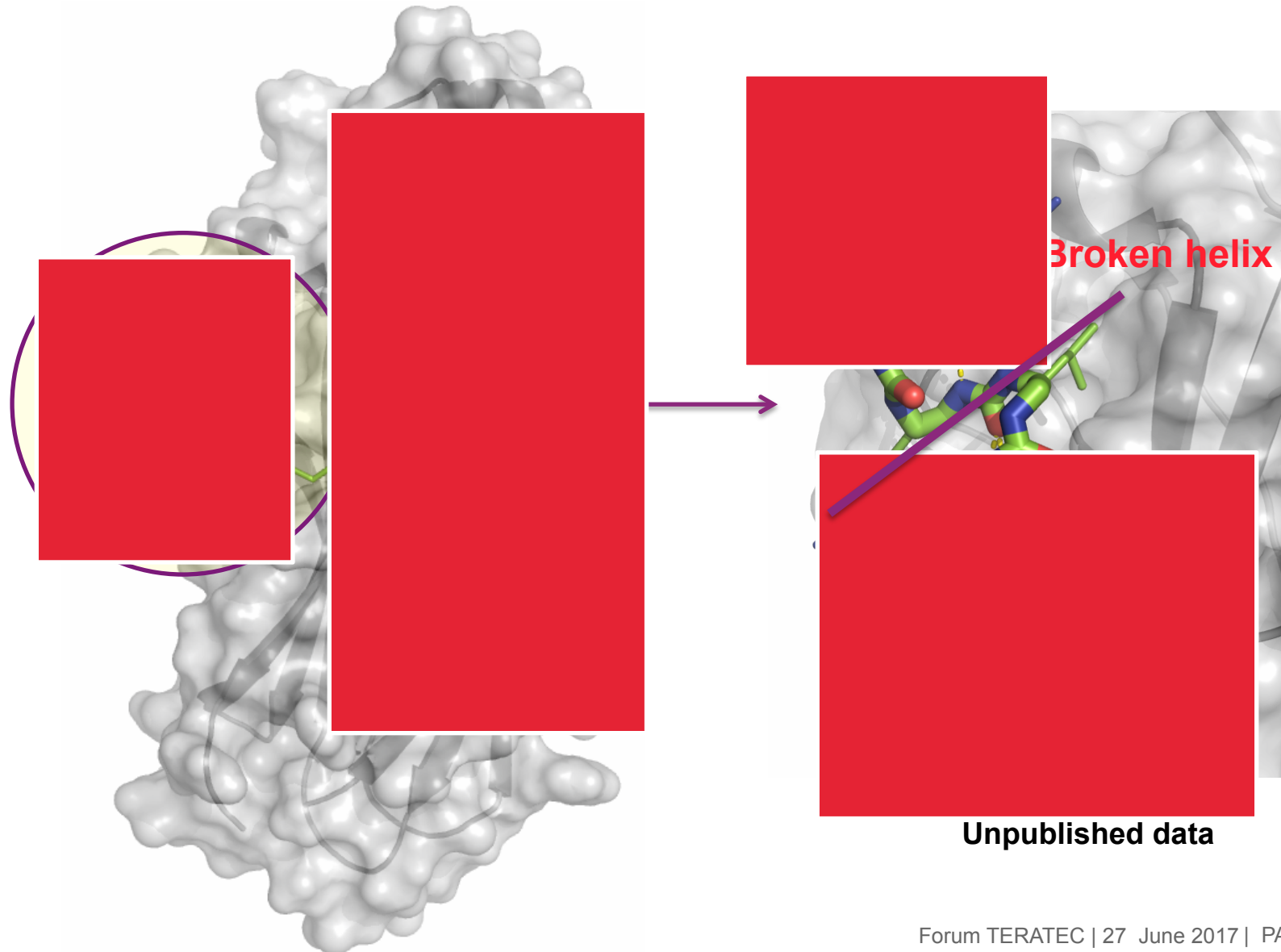
Xray

(1.9 Å)

3nM

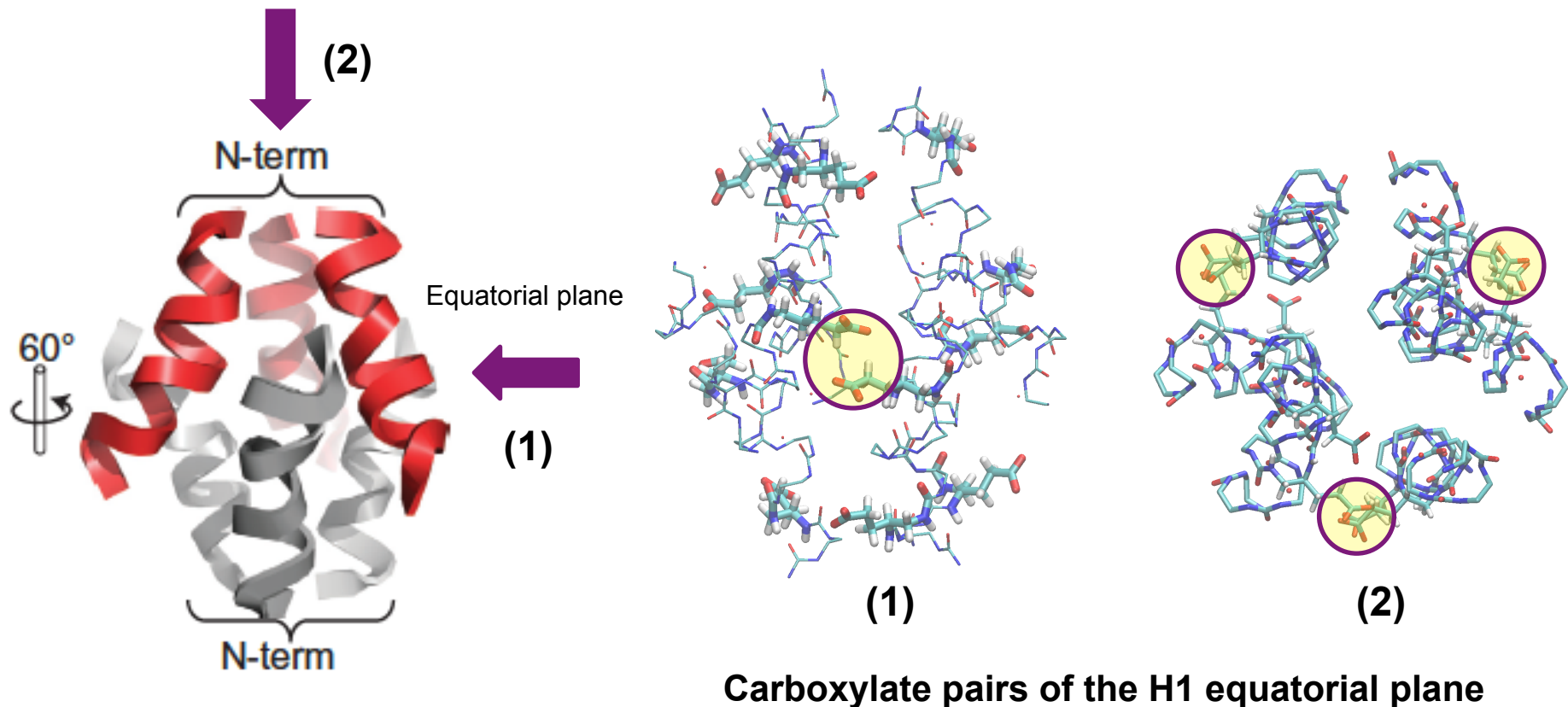Peptide ████ ligand developed by the ████████████████ teams

Broken helix

**Unpublished data**

**2x10 MD simulations of the solvated H1 hexamer (1 solvent level, 5 ns) for two carboxylate protonation states (fully unprotonated, or half of the carboxylates of the H1 equatorial plane protonated)**



**Carboxylate pairs of the H1 equatorial plane**

→ **All the simulations converged towards the same hexamer structure in aqueous phase, regardless of the carboxylate protonation states**
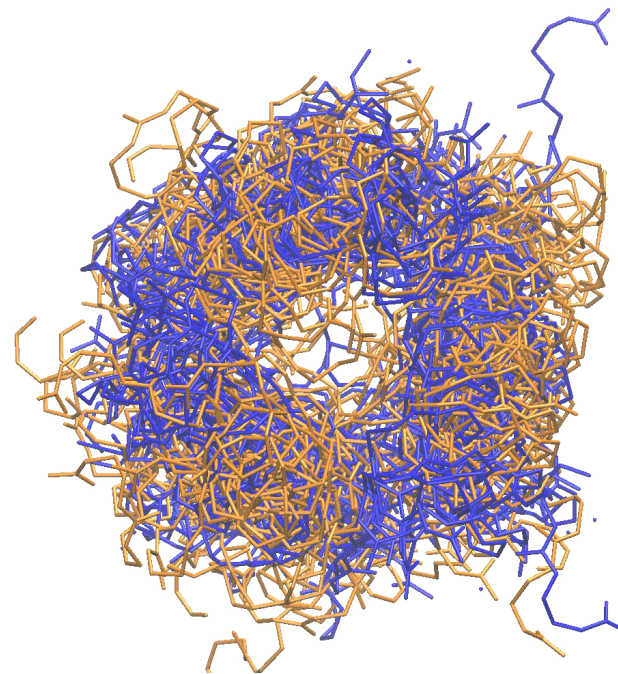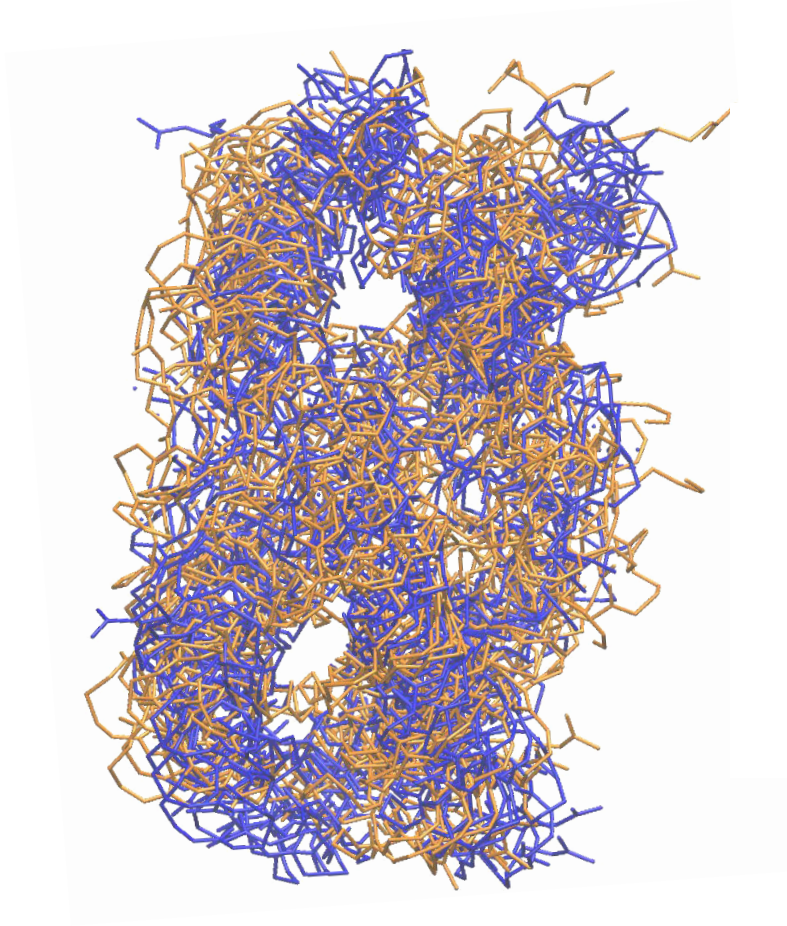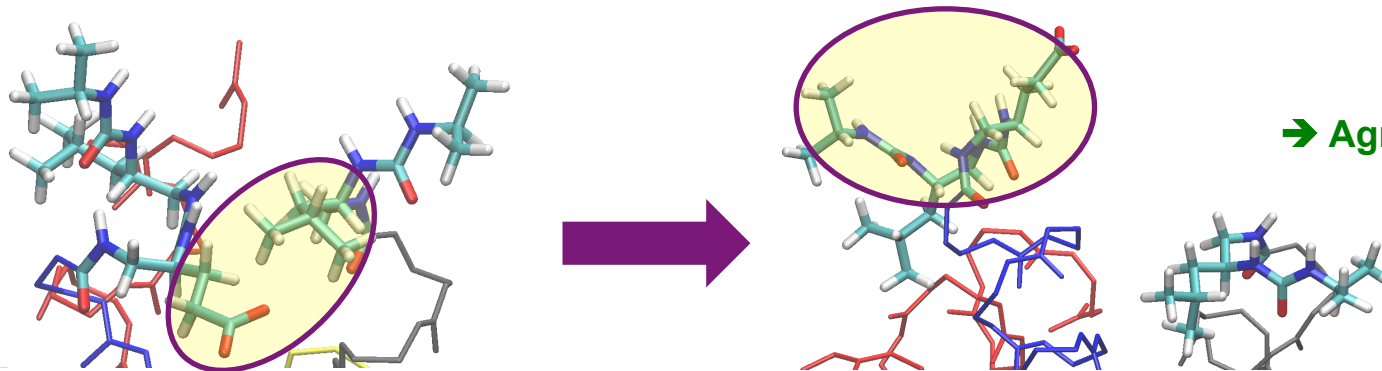
**Tyr\*5/Leu\*11 : intermolecular interaction observed along the trajectories**
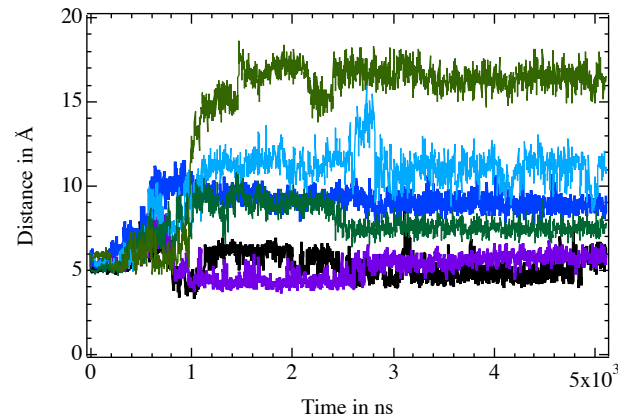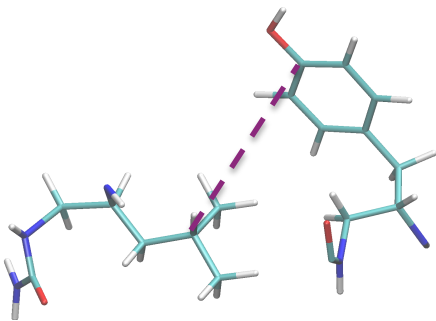


➔ **Agree with experiment**

**Leu\*1/Glu\*2 : intermolecular interaction disrupted ➔ intramolecular interaction**



➔ **Agree with experiment**

27 JUNE 2017

R 2012

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

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

Life Science Division
IBITECS/SIMOPRO
Life chemistry laboratory