# Extending the MST model to large biomolecular systems: parametrization of the ddCOSMO-MST continuum solvation model

UNIVERSITAT DE BARCELONA

**Renato D. Cunha, PhD**
*Postdoctoral Researcher*
*Computational Photobiology Lab*
*Facultad de Farmàcia i Ciències de l'Alimentació*

**Cunha, R. D.*, Romero, S., Lipparini, F., Luque, F. J., Curutchet, C. (2025). *Extending the MST model to large biomolecular systems: parametrization of the ddCOSMO-MST continuum solvation model. Journal of Computational Chemistry. 46(2),* e70027.

# Brief presentation of the Miertus-Scrocco-Tomasi (MST)

The MST continuum solvation method uses a dual cavity strategy, defining different solute/solvent interfaces for the electrostatic and nonelectrostatic components.

**Electrostatic Contribution**: This term accounts for the work required to charge the solute in the solvent compared to a vacuum. The solvent is treated as a continuous polarizable medium that generates a reaction field in response to the solute's charge distribution.

**Cavitation Contribution:** This term represents the work needed to create a cavity in the solvent to accommodate the solute. It is computed using Pierotti's scaled particle theory, adapted to molecular-shaped cavities.

**Van der Waals Contribution:** This term accounts for dispersion-repulsion interactions between solute and solvent molecules. It is determined using a linear relationship between the van der Waals free energy of an atom and its solvent-exposed surface area.

$$\Delta G_{\mathrm{sol}} = \Delta G_{\mathrm{ele}} + \Delta G_{\mathrm{n-ele}}$$

$$\Delta G_{\mathrm{n-ele}} = \Delta G_{\mathrm{cav}} + \Delta G_{\mathrm{vW}} = \sum_i \frac{S_i}{S_T} \Delta G_{P,i} + \sum_i \xi_i S_i$$

*C. Curutchet, M. Orozco, and F. J. Luque. (2001) Journal of Computational Chemistry. 22(11). 1180–1193.*
*C. Curutchet, A. Bidon-Chanal, I. Soteras, M. Orozco, and F. J. Luque. (2005). Journal of Physical Chemistry B. 109(8). 3565–3574*

# Brief presentation of the ddCOSMO

A scalable implementation of the COSMO model for implicit solvation in quantum chemistry.

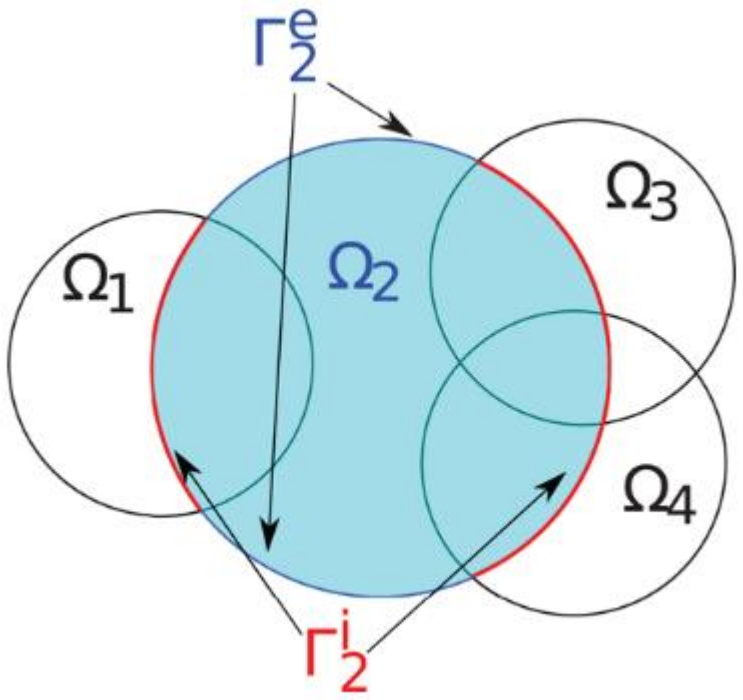
The molecular surface is divided into small, overlapping regions (domains).

Achieves linear scaling for large molecules.

Developed to address the computational limitations of traditional COSMO for large systems.

Each domain is processed independently and then combined to construct the full solvation potential.

Enables parallel processing for efficiency on modern computational architectures



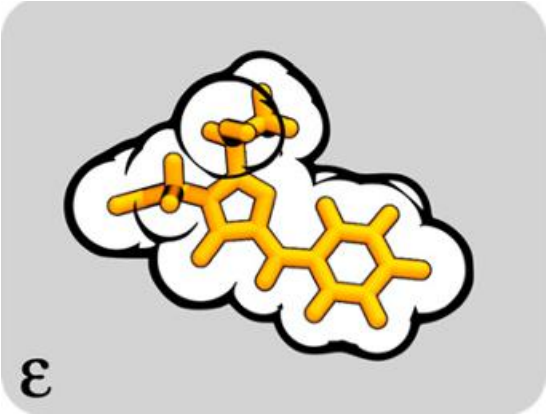

*Lipparini, F., Stamm, B., Cancès, E., Maday, Y., & Mennucci, B. (2013). Journal of Chemical Theory and Computation, 9(8), 3637–3648.*

Parametrization of ddCOSMO to predict hydration free energies

We introduced several changes in the MST parametrization strategy

Speed up calculations by three orders of magnitude

MST adopt a solvent-specific cavity scaling, applied to the radii to define solute cavity



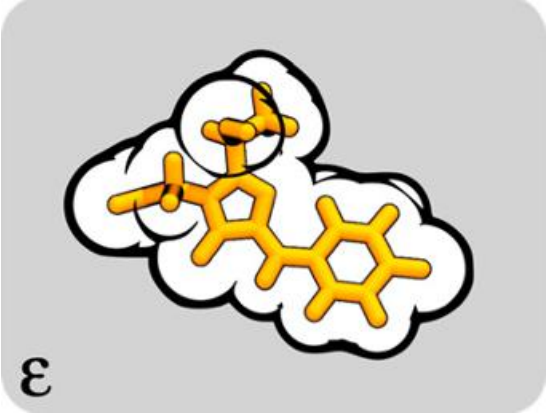

New definition of atom types for non-electrostatic contributions based on hybridization aimed at a larger transferability

Automatized and smooth adaptation of the cavity size in charged regions, to facilitate tackling large biomolecules

Parametrization of ddCOSMO to predict hydration free energies

Speed up calculations by three orders of magnitude

MST adopt a solvent-specific cavity scaling, applied to the radii to define solute cavity




We introduced several changes in the MST parametrization strategy

New definition of atom types for non-electrostatic contributions based on hybridization aimed at a larger

he cavity size ackling large

We implemented a two-step SCF process, in which we first converge the SCF equations with the "neutral" cavity obtained by applying only $\lambda_{solv}$. This is used to estimate $\langle V_i \rangle$, which is a useful index for chemical reactivity sensible to the influence of external fields. Then, the resulting $\lambda_i$ parameters are used to build the refined cavity used to reconverge the definitive SCF equations.

$$\lambda_i(\langle V_i \rangle) = \lambda_{\text{solv}} \left[ \frac{1+\alpha}{2} - \frac{1-\alpha}{2} \tanh\left( \frac{|\langle V_i \rangle \cdot R_i| - V_{\text{mid}}}{W} \right) \right]$$

# Computational details

$$\Delta G_{sol}^{res} = \Delta G_{sol}^{exp} - \Delta G_{ele} - \Delta G_{cav}$$

**Training set**

Parametrizantion of the van der Waals surface tensions $\xi_i$

Training set of 229 neutral molecules in aqueous solution

Performed a multiple linear regression of the residual term ($\Delta G_{sol}^{res}$)

Optimization of the parameters $\alpha$, $V_{mid}$ and $W$ for ionic training

Training set of 122 molecules of the Minnesota Solvation Database*

Optimization: PM6 and B3LYP/6-31G(d) levels of theory using ddCOSMO

**Test set**

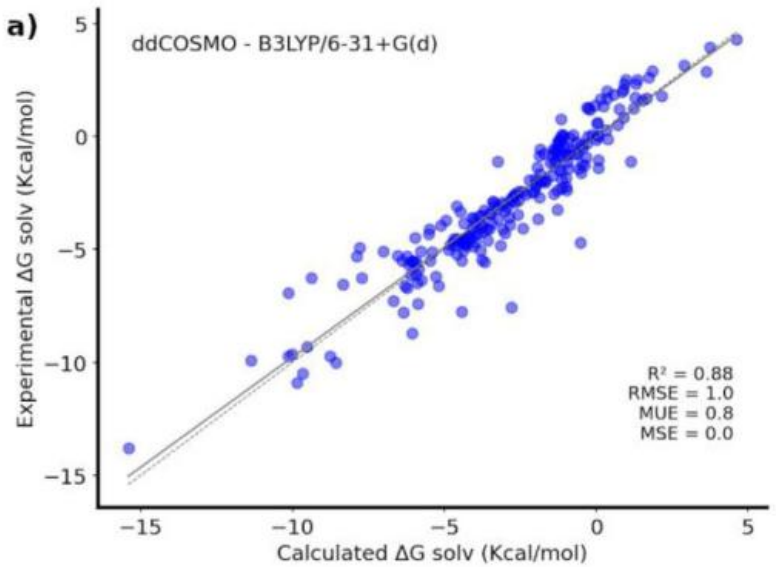
Applied on the test set using the drug-like molecules by SAMPL2 and SAMPL4 challenge

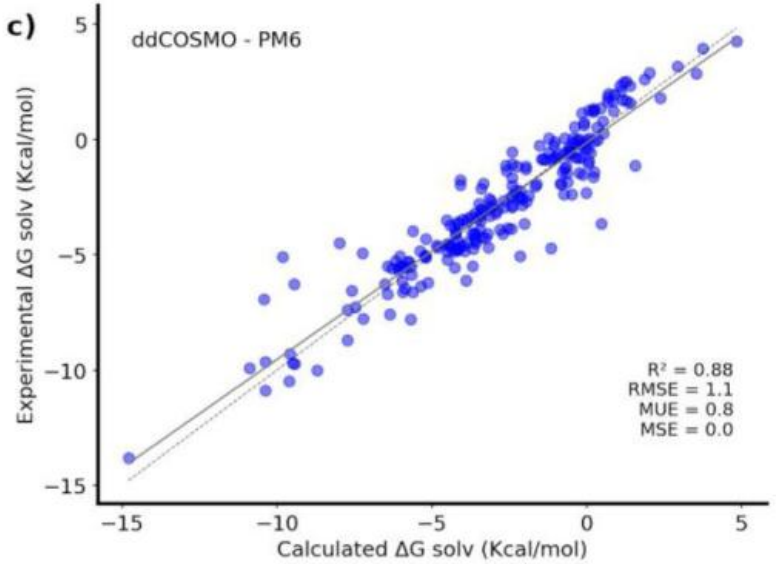
Solvation free energies computed in vacuum and in water solution at the PM6 and B3LYP/6-31+G(d)

*Marenich, Aleksandr V; Truhlar, D. G. et al. (2020). Retrieved from the Data Repository for the University of Minnesota (DRUM).*
*Skillman, A. G., et al. (2010). Journal of Computer-Aided Molecular Design, 24(4), 257–258.*
*Muddana, H. S.,et al. (2014).. Journal of Computer-Aided Molecular Design, 28(4), 305–317.*

# Training set – Neutrals

| | ddCOSMO/MST | | IEFPCM/MST | |
|---|---|---|---|---|
| | B3LYP/6-31+G(d) | PM6 | B3LYP/6-31+G(d) | PM6 |
| **MSE** | 0.0 | 0.0 | 0.0 | 0.0 |
| **MUE** | 0.8 | 0.8 | 0.8 | 0.7 |
| **RMSE** | 1.0 | 1.1 | 1.0 | 1.0 |

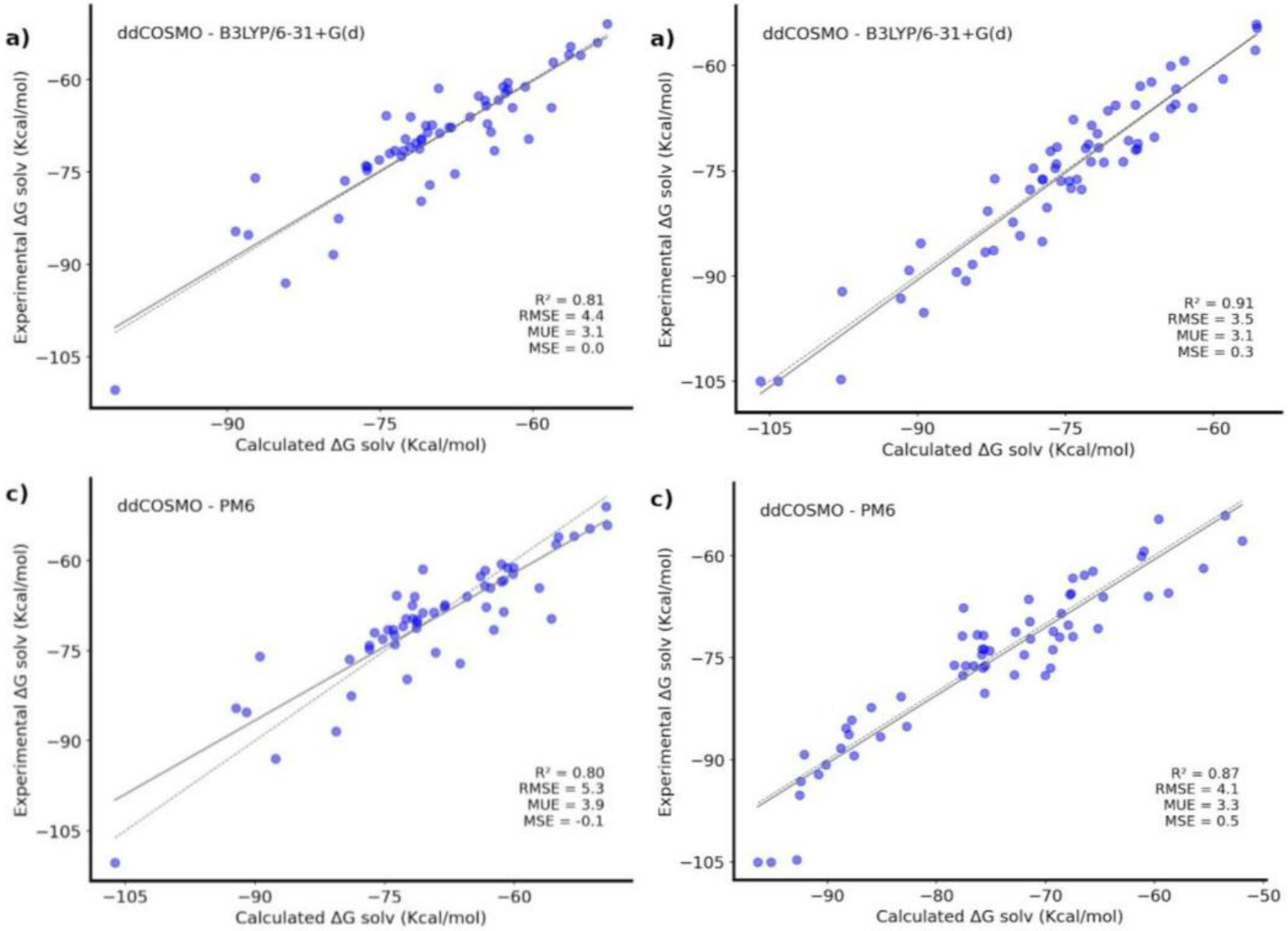


# Training set – Ions

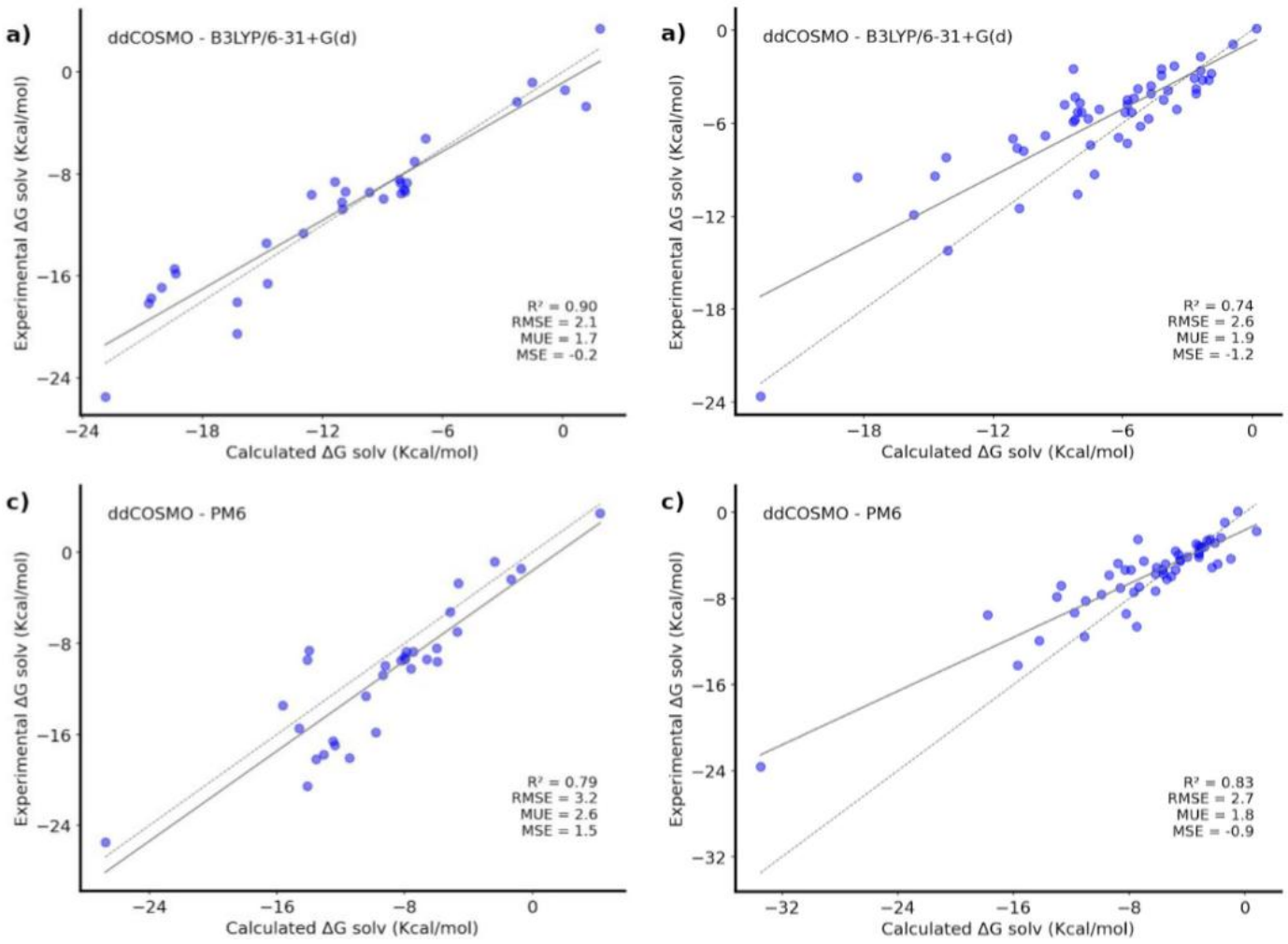
| MUE | Cations | Anions |
|---|---|---|
| **ddCOSMO** | | |
| **B3LYP/6-31G(d)** | 3.2 | 4.1 |
| **B3LYP/6-31+G(d)** | 3.1 | 3.1 |
| **PM6** | 3.9 | 3.3 |
| **IEFPCM** | | |
| **B3LYP/6-31G(d)** | 3.4 | 3.8 |
| **B3LYP/6-31+G(d)** | 3.4 | 5.0 |
| **PM6** | 4.1 | 3.1 |

IEFPCM suffers more outlying charge problem in anions



# Test set

86 small and drug-like molecules, 40 and 46 from SAMPL2 and SAMPL4, respectively, and 10 charged molecules of the C10 dataset;



# Test set

| | SAMPL2 | | | SAMPL4 | | | C10[a] | | |
|---|---|---|---|---|---|---|---|---|---|
| | MSE | MUE | RMSE | MSE | MUE | RMSE | MSE | MUE | RMSE |
| ddCOSMO | | | | | | | | | |
| B3LYP/6-31+G(d) | −0.2 | 1.7 | 2.1 | −1.2 | 1.9 | 2.6 | 0.4 | 1.6 | 2.2 |
| | | | | | | | (0.3) | (1.6) | (2.3) |
| PM6 | 1.5 | 2.6 | 3.2 | −0.9 | 1.8 | 2.7 | 0.2 | 1.9 | 2.4 |
| | | | | | | | (0.1) | (1.4) | (2.1) |
| IEFPCM | | | | | | | | | |
| B3LYP/6-31+G(d) | −0.3 | 1.5 | 1.9 | −1.0 | 1.8 | 2.4 | 1.3 | 3.6 | 4.7 |
| | | | | | | | (1.2) | (3.7) | (4.7) |
| PM6 | 1.8 | 2.9 | 3.6 | −0.5 | 1.6 | 2.3 | −1.0 | 2.6 | 3.2 |
| | | | | | | | (−1.3) | (2.1) | (2.6) |

*Note:* All values in kcal/mol.
[a]Values in parenthesis are determined using the geometries provided for the C10 dataset.

# Conclusions

We reported the parametrization of the ddCOSMO-MST and IEFPCM-MST continuum models for the prediction of hydration free energies of neutral and ionic compounds at the B3LYP/6-31+G(d), B3LYP/6-31G(d) and PM6.

Beyond the novel ddCOSMO electrostatic formalism, we introduced other modifications in MST like a definition of atom types for non-electrostatic contributions based on hybridization state and a simple yet effective automatic setup of the cavity size for charged regions based on sensing the electrostatic potential at the cavity surface.

The final ddCOSMO/MST models provide excellent mean unsigned errors ~0.7-0.8 kcal/mol for neutrals and ~3.1-3.4 kcal/mol for ions both for the DFT and PM6 models, with similar or improved accuracies compared to IEFPCM-MST at a reduced computational cost.

*C. Curutchet, M. Orozco, and F. J. Luque. (2001) Journal of Computational Chemistry. 22(11). 1180–1193.*
*C. Curutchet, A. Bidon-Chanal, I. Soteras, M. Orozco, and F. J. Luque. (2005). Journal of Physical Chemistry B. 109(8). 3565–3574*

# THANK YOU!

**Computational Photobiology Lab**

**PI -** Dr. Carles Curutchet
Dr. Anton J. Stasyuk
Dr. Renato D. Cunha
Daniel Gonzalo
Beste Ozaydin
Özge Ergun
Blanca Díaz

**Collaborators**
Dr. F. Javier Luque
Dr. Filippo Lipparini

**Fellowship:** PREDOCS-UB2020 (Postdoctoral Period – POP)