Towards accurate solvation free energies of large biological systems

S. Romero,[a] F. J. Luque,[a] X. Barril,[a] F. Lipparini,[b] B. Mennucci,[c] C. Curutchet,[a]

[a]Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Spain
[b]Laboratoire Jacques-Louis Lions - Laboratoire de Chimie Theórique - Institut du Calcul et de la Simulation, Sorbonne Universiteś, Paris, France
[c]Dipartamento di Chimica e Chimica Industriale, Università di Pisa, Italy.
E-mail: sonia.romero@ub.edu

Abstract- Continuum solvation models like PCM or COSMO are the standard tool to calculate solvation free energies in a quantum level, but have been typically limited to small biological molecules due to its large computational cost. Recently, a new implementation of COSMO based on a domain decomposition strategy (ddCOSMO) [1] has been presented, which speeds up calculations by several orders of magnitude, thus paving the way for its application to very large systems. Here, we report the parameterization of ddCOSMO to the prediction of hydration free energies based on the MST solvation model developed in Barcelona, [2][3]. The parameterization is based on the PM6 semi-empirical Hamiltonian, on a set of over 200 experimental hydration free energies. The new model opens the way to the accurate prediction of hydration free energies of very large biomolecules, thus going beyond the usual classical MM-PBSA or MM-GBSA approaches.

Keywords: Implicit solvation models, MST Solvation Model, ddCOSMO.

I. INTRODUCTION

In continuum solvation models, the solute is treated at a QM or MM level, and the solvent is described as a continuum dielectric medium.

The solvation free energy is then computed as a sum of electrostatic , cavitation and Van der Waals free energies, where the last two are defined as non-electrostatic term.

$$\Delta G_{solv} = \Delta G_{ele} + \Delta G_{noele} \qquad (1)$$

Where the electrostatic term is defined as:

$$\Delta G_{ele} = \langle \Psi^{solv}|H^0 + {}^1\!/_2 V^{solv}|\Psi^{solv}\rangle - \langle \Psi^0|H^0|\Psi^0\rangle \qquad (2)$$

There are different strategies to solve the electrostatic problem such as Generalized Born and Poisson-Boltzmann methods, which are used mostly in Molecular Mechanics implementations.

In quantum mechanics implementations, Apparent surface charge methods such as Polarizable continuum model (PCM) and Conductor-like screening models (c-PCM or COSMO) are chosen. Nevertheless, its computational cost only allows using these models in small systems.

ddCOSMO is a recently proposed algorithm to solve the polarization equation for the Conductor-like Screening Model (COSMO, where the electrostatic solute-solvent interaction energy is obtained as:

$$E_{ele} = {}^1\!/_2 \, f(\varepsilon) \int_\Omega \rho(\mathbf{r}) \, W(\mathbf{r}) d\mathbf{r} \qquad (3)$$

Where f(ε) is an empirical scaling introduced to account for the non-conductor nature of the solvent and ε is its dielectric constant, ρ is the charge density of the solute and W is the polarization potential W of the conductor, usually referred to as the reaction field.

The ddCOSMO model[1] solves the COSMO equations based on Schwarz's domain decomposition method, and has been proven to be both smooth and fast; furthermore, linear scaling in both computational cost and memory requirements with respect to the system's size is implicit in the procedure without needing to resort to fast summation techniques. With respect to existing linear-scaling implementations, ddCOSMO can be two to three orders of magnitude faster, allowing computing the solvation energy for very large systems with a reduced computational cost.

In this project, we are re-parameterizing MST solvation model using ddCOSMO algorithm, at B3LYP and more recent PM6 semi-empirical level on a set of over 200 neutral molecules. The aim is to obtain free solvation energies at a quantum level even of large biological systems, in a cheaper and faster way.

II. COMPUTATIONAL DETAILS

The training set is of 238 neutral molecules with known experimental solvation free energies from Cramer and Truhlar data set [3]. All molecules have been optimized in gas phase and solution, (parameterizations are performed for both sets of geometries). The electrostatic free energy is easily computed using MST cavity settings and ddCOSMO method. The non-electrostatic term is isolated in equation (1) and computed using experimental solvation free energies. Then fitted using a multiple linear regression method.

$$\Delta G_{noele} = \sum_{i=1}^{N} \xi_i S_i \qquad (4)$$

The non-electrostatic free energy for each molecule will be obtained multiplying the atomic surface tensors (ξ_i) by the surface of each element/hybridization atom type.

III. RESULTS AND DISCUSSION

Our calculated solvation free energies are compared with the experimental ones, obtaining good results for both PM6 and B3LYP theory levels, being B3LYP slightly better. These results were independently of which geometry optimization phase we

used.
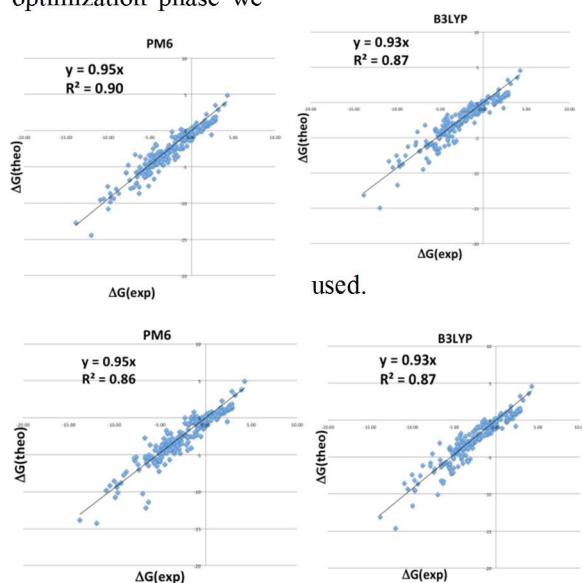
Fig. 1. Comparison between experimental and calculated ΔG_{solv}. Parameterization of **a:** PM6 level, molecules optimized in gas phase. **b**: B3LYP level, molecules optimized in gas phase. **c:** PM6 level, optimization in solution. d: B3LYP level, optimization in solution

In the following table, are described the Mean Signed Error (MSE), Mean Unsigned Error (MUE)

	MSE	MUE	RMSD
PM6 gas	0,05	0,74	0,97
B3LYP gas	0,03	0,83	1,08
PM6 solv	0,04	0,86	1,14
B3LYP solv	0,03	0,83	1,07

and Root Mean Squared Deviation (RMSD) of each parameterization.

Two different atom type sets were used in the parameterization: i) Element atom type (9), used in MST Model, define an atom type for each element (H, C, O, N,S , F, Cl, Br, P) and ii) Hybridization type(15), (H, Hp, Csp, Csp2, Csp3, Osp2, Osp3, Nsp, Nsp2, Nsp3,S , F, Cl, Br, P). This second atom type definition was definitely better than the element type.

IV. CONCLUSIONS AND FUTURE PERSPECTIVES

Both PM6 and B3LYP parameterizations are able to accurately describe the experimental hydration free energies of neutral molecules with errors below 1 kcal/mol.

Future work will extend the parameterization to charged molecules, based on an automatic rescaling of the cavity size in charged regions, following previous work in the context of the MST model.

REFERENCES

[1] M.Orozco,F.J.Luque, "Theoretical Methods for the Description of the Solvent Effect in Biomolecular Systems", Chem.Rev.100(2000), 4187.

[2]F.Lipparini,L.Lagardère,G.Scalmani,B.Stamm,E.Cancès,Y.M aday,J.-P.Piquemal,M.J.Frisch,B.Mennucci, "Quantum Calculations in Solution for Large to Very Large Molecules: A New Linear Scaling QM/Continuum Approach" J.Phys.Chem.Lett.(2014)953.

[3]C.Curutchet,M.Orozco,F.J.Luque,J.Comput.Chem.22(2001), 1180;C.Curutchet,A.BidonChanal,I.Soteras,M.Orozco,F.J. Luque,J.Phys.Chem.B109(2005),3565.[4]A.V.Marenich,C. J.Cramer,D.G.Truhlar,J.Phys.Chem.B113(2009),6378

[4]A.V.Marenich,C.J.Cramer,D.G.Truhlar,J.Phys.Chem.B113(2 009),6378.