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

R. D. Cunha[1,2]*, S. Romero[3], F. Lipparini[4], F. J. Luque[2,5,6], C. Curutchet[1,2]

[1]*Departament de Farmàcia i Tecnologia Farmacèutica, i Fisicoquímica, Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona (UB), Barcelona, Spain*
[2]*Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona (UB), Barcelona, Spain*
[3]*Departament de Química and Institut de Biotecnologia i de Biomedicina (IBB), Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain*
[4]*Dipartimento di Chimica e Chimica Industriale, Università di Pisa, Pisa, Italy*
[5]*Departament de Nutrició, Ciències de l'Alimentació i Gastronomia, Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona (UB), Santa Coloma de Gramenet, Spain*
[6]*Institut de Biomedicina (IBUB), Universitat de Barcelona (UB), Barcelona, Spain*
**Presenting author e-mail address: renatodias@ub.edu*

ddCOSMO, MST continuum solvation model, DFT, semiempirical methods

INTRODUCTION

Continuum solvation models are popular to describe solvation effects in computational chemistry and are widely applied to study either small molecules and large biomolecular systems. In the conductor-like screening model (COSMO) the differential equations are recast as integral equations on the cavity boundary, which can be treated with standard numerical techniques such as the boundary element method (BEM) by discretizing the cavity surface using a mesh. Recently, a different strategy to solve the COSMO problem has been presented based on Schwarz's domain decomposition strategy (ddCOSMO).[1] The novel linear-scaling formulation speeds up calculations by several orders of magnitude, thus paving the way for its application to large biomolecular systems.

Here, we report the parametrization of ddCOSMO to the prediction of hydration free energies based on the MST solvation model developed in Barcelona,[2][3] which has been previously successfully applied to a variety of biological problems, [4][5] e.g. tautomeric equilibria, octanol-water partition coefficients, acidity constants, or the definition of 3D hydrophobic/hydrophilic profiles of biomolecules. Beyond ddCOSMO electrostatics, we introduce several novelties in MST, like a new definition of atom types for non-electrostatic contributions based on hybridisation and an automatic adaptation of the cavity size for charged regions of ions.

METHODS

In the MST quantum chemical continuum model, the solvation free energy (ΔGsol) includes electrostatic (ΔGele) and non-electrostatic (ΔGnon-ele) contributions:

$$\Delta G_{sol} = \Delta G_{ele} + \Delta G_{no-ele}$$

ΔGele is computed using the ddCOSMO implementation in a local development versión of Gaussian, and the ΔGnon-ele term is computed as a cavitation term (using Claverie-Pierotti theory) and a surface area term, accounting for disperison-repulsión interactions:

$$\Delta G_{no-ele} = \Delta G_{vdW} + \Delta G_{cav} = \sum_i \xi_i S_i + \sum_i \frac{S_i}{S_T} \Delta G_{P,i}$$

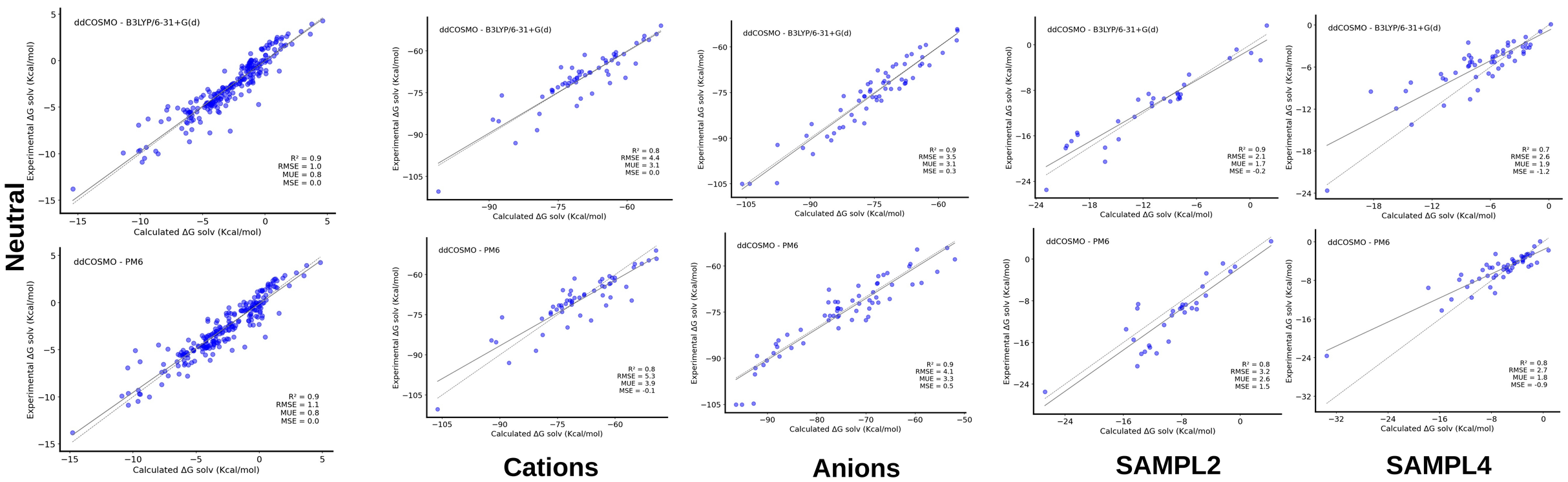
Where *Si* is the area of each atom, and ξ the atomic surface tensions.

The cavity is automatically adjusted in charged regions using the average QM potential in the atomic surface, scaling atomic radii by this function:

$$f_i(\langle V_i \rangle) = \frac{1+\alpha}{2} - \frac{1-\alpha}{2} tanh\left(\frac{\left|\langle V_i \rangle \cdot R_i\right| - V_{mid}}{W}\right)$$

Training set: Neutrals (229 molecules), Ions (58 cations and 64 anions)
- Test set: SAMPL2 (40 molecules), SAMPL4 (52 molecules)
- B3LYP/6-31+G(d) and semi-empirical PM6 levels of theory.

RESULTS



Atomic surface tensions (kcal/mol·Å²)

Atom type	B3LYP/6-31+G(d)	PM6
H	-0.1238	-0.1193
Hp	0.1126	-0.0141
Csp	-0.0578	-0.1270
Csp2	-0.1409	-0.1492
Csp3	-0.2003	-0.1764
Nsp	-0.0368	-0.0428
Nsp2	-0.2162	-0.1442
Nsp3	-0.2623	-0.1580
Osp2	-0.0676	-0.0485
Osp3	-0.1568	-0.1767
F	-0.0618	-0.0804
P	-0.1942	-0.1334
S	-0.1104	-0.0578
Cl	-0.1165	-0.1090
Br	-0.1250	-0.0982

Parameters for scaling function

	α	V_{mid} (a. u.)	W (a. u.)
B3LYP/6-31+G(d)			
Cations	0.94	0.40	0.20
Anions	0.84	0.50	0.20
PM6			
Cations	0.92	0.40	0.10
Anions	1	--	--

CONCLUSION

The new strategy to solve the COSMO equations (ddCOSMO) allows to tackle solvation energy calculations of large molecules, and this work presents the parametrization of non-electriostatic contributions based on the MST model. The results show a similar performance of MST-ddCOSMO compared to more rigorous methods like MST-IEFPCM and other solvation models reported in previous SAMPL2 and SAMPL4 challenges. Future extensions of the method will include other solvents, for instance n-octanol, to perform log P calculations of drug-like molecules, and the parametrization of ddCOSMO at the polarizable and non-polarizable MM force field level, to allow for molecular dynamics simulations of biomolecules in implicit solvent.

ACKNOWLEDGEMENT

CEX-2021-001202-M, PID2023-151584NB-100

REFERENCES

[1] F. Lipparini, L. Lagardère, G. Scalmani, B. Stamm, E. Cancès, Y. Maday, J.-P. Piquemal, M. J. Frisch, B. Mennucci, J. Phys. Chem. Lett. 5 (2014) 953.
[2] C. Curutchet, M. Orozco, F.J. Luque, J. Comput. Chem. 22 (2001), 1180.
[3] C. Curutchet, A. Bidon-Chanal, I. Soteras, M. Orozco, F.J. Luque, J. Phys. Chem. B 109 (2005), 3565.
[4] W. J. Zamora, S. Pinheiro, K. German, C. Ràfols, C. Curutchet and F. J. Luque, J. Comput. Aid. Mol. Des. 2020, 34, 443-451.
[5] A. Viayna, S. Pinheiro, C. Curutchet, F. J. Luque, W. J. Zamora, J. Comput. Aid. Mol. Des. 2021, 35, 803-811.