

# Quantum calculations in solution for large to very large molecules: a new linear scaling QM/continuum approach

Filippo Lipparini,<sup>\*,†,‡,¶</sup> Louis Lagardère,<sup>¶</sup> Giovanni Scalmani,<sup>§</sup> Benjamin Stamm,<sup>†,||</sup>  
Eric Cancès,<sup>⊥</sup> Yvon Maday,<sup>†,#,@</sup> Jean-Philip Piquemal,<sup>‡,||</sup> Michael J. Frisch,<sup>§</sup> and  
Benedetta Mennucci<sup>△</sup>

*Sorbonne Universités, UPMC Univ. Paris 06, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France, Sorbonne Universités, UPMC Univ. Paris 06, UMR 7616, Laboratoire de Chimie Théorique, F-75005, Paris, France, Sorbonne Universités, UPMC Univ. Paris 06, Institut du Calcul et de la Simulation, F-75005, Paris, France, Gaussian, Inc., 340 Quinpiac St. Bldg. 40, Wallingford, Connecticut 06492, USA., CNRS, UMR 7598 and 7616, F-75005, Paris, France, Université Paris-Est, CERMICS, Ecole des Ponts and INRIA, 6 & 8 avenue Blaise Pascal, 77455 Marne-la-Vallée Cedex 2, France, Institut Universitaire de France, Brown Univ, Division of Applied Maths, Providence, RI, USA, and Dipartimento di Chimica e Chimica Industriale, Università di Pisa, Via Risorgimento 35, 56126 Pisa, Italy*

E-mail: [filippo.lipparini@courriel.upmc.fr](mailto:filippo.lipparini@courriel.upmc.fr)

In this supporting information, we will briefly derive the ddCOSMO contributions to the Fock matrix for a semiempirical Hamiltonian. Let  $\rho$  be the density of charge of the solute. The COSMO solves the electrostatics equation for such a density, accommodated into a hollow cavity  $\Omega$  embedded into a metallic, infinite continuum. Here, the cavity is assumed to be the union of intersecting spheres, which is the case for both Van der Waals and Solvent Accessible Surface (SAS) cavities. Such a problem is solved by writing the total electrostatic potential as the sum of two terms: the potential produced by  $\rho$  *in vacuo* and a reaction contribution due to the polarization of the metal, usually called  $W$ . The electrostatic solvation energy is one half of the interaction energy between the inducing energy  $\rho$  and the reaction potential  $W$ :

$$E_s = \frac{1}{2}f(\varepsilon) \int_{\Omega} \rho(\mathbf{r})W(\mathbf{r})d\mathbf{r} \quad (1)$$

In the ddCOSMO paradigm, the reaction potential is computed by solving a collection of interacting problems, one for each sphere composing the cavity, in an iterative fashion. In particular, for each sphere  $\Omega_j$  the reaction potential can be expressed as

$$\forall \mathbf{r} \in \Omega_j, \quad W(\mathbf{r}) = \int_{\Gamma_j} \frac{\sigma_j(\mathbf{s})}{|\mathbf{r} - \mathbf{s}|} d\mathbf{s} := (\mathcal{S}_j \sigma_j)(\mathbf{r}) \quad (2)$$

---

\*To whom correspondence should be addressed

<sup>†</sup>Sorbonne Universités, UPMC Univ. Paris 06, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France

<sup>‡</sup>Sorbonne Universités, UPMC Univ. Paris 06, UMR 7616, Laboratoire de Chimie Théorique, F-75005, Paris, France

<sup>¶</sup>Sorbonne Universités, UPMC Univ. Paris 06, Institut du Calcul et de la Simulation, F-75005, Paris, France

<sup>§</sup>Gaussian, Inc., 340 Quinpiac St. Bldg. 40, Wallingford, Connecticut 06492, USA.

<sup>||</sup>CNRS, UMR 7598 and 7616, F-75005, Paris, France

<sup>⊥</sup>Université Paris-Est, CERMICS, Ecole des Ponts and INRIA, 6 & 8 avenue Blaise Pascal, 77455 Marne-la-Vallée Cedex 2, France

<sup>#</sup>Institut Universitaire de France

<sup>@</sup>Brown Univ, Division of Applied Maths, Providence, RI, USA

<sup>△</sup>Dipartimento di Chimica e Chimica Industriale, Università di Pisa, Via Risorgimento 35, 56126 Pisa, Italy

where  $\Gamma_j = \partial\Omega_j$  and the local apparent surface charge (ASC)

$$\sigma_j = \sum_{l=0}^N \sum_{m=-l}^l [\sigma_j]_l^m Y_l^m(\mathbf{s})$$

is an intermediate quantity used in the ddCOSMO algorithm, which is obtained by solving the ddCOSMO linear equations  $L[\sigma] = g$ . All the details can be found elsewhere;<sup>1,2</sup> here it is sufficient to say that the right-hand side  $g$  is the solute's electrostatic potential (and therefore a quantity linear in the density matrix) weighted with a switching function and that the  $L$  matrix is not symmetric, but very sparse.

In order to couple the ddCOSMO to a SCF-like procedure, the ddCOSMO Fock matrix contribution has to be computed. Such a contribution is defined as the derivative of the ddCOSMO solvation energy with respect to the electronic density matrix  $P_{\mu\nu}$ :

$$F_{\mu\nu}^s = \frac{\partial E_s}{\partial P_{\mu\nu}} = \frac{1}{2}f(\varepsilon) \int_{\Omega} \chi_{\mu}(\mathbf{r})\chi_{\nu}(\mathbf{r})W(\mathbf{r}) + \rho(\mathbf{r})\frac{\partial W(\mathbf{r})}{\partial P_{\mu\nu}}d\mathbf{r}, \quad (3)$$

where the electronic density of the solute has been expanded in a basis of AO:

$$\rho(\mathbf{r}) = \sum_{\mu\nu} P_{\mu\nu}\chi_{\mu}(\mathbf{r})\chi_{\nu}(\mathbf{r}).$$

For semiempirical methods, both equations 1 and 3 simplify greatly as the electrostatic properties can be described as the ones produced by a collection of atomic monopoles, dipoles and quadrupoles, which are related to the density matrix through a linear transformation, which depends on the specific parametrization. Let  $[\Theta_j]_l^m$ ,  $0 \leq l \leq 2$ ,  $-l \leq m \leq l$  be such atomic multipoles (normalized so that the monopole is actually a charge, the  $l = 1$  terms correspond to a dipole and the  $l = 2$  to a quadrupole). The integral in eq. 1 simplifies to

$$E^s(\Theta) = \frac{1}{2}f(\varepsilon) \sum_{j=1}^M \int_{\Omega_j} d\mathbf{r} \sum_{l=0}^2 \sum_{m=-l}^l \sqrt{\frac{2l+1}{4\pi}} [\Theta_j]_l^m \delta(r) Y_l^m(\theta, \phi) W(r, \theta, \phi) \quad (4)$$

By exploiting equation 2 and using a multipolar expansion of the integral there involved, eq. 5 is recovered:

$$E^s(\boldsymbol{\Theta}) = \frac{1}{2}f(\varepsilon) \sum_{j=1}^M \sum_{l=0}^2 \sum_{m=-l}^l \sqrt{\frac{4\pi}{2l+1}} [\Theta_j]_l^m [\sigma_j]_l^m. \quad (5)$$

Now, for each atom  $i$  and for each atomic multipole  $l, m$ , let

$$[\Theta_i]_l^m = \sum_{\mu(i)\nu(i)} \Lambda_{ilm,i\mu\nu} P_{\mu\nu} \quad (6)$$

where the indexes  $\mu(i), \nu(i)$  correspond to basis functions centered on the  $i$ -th atom. Notice that the matrix  $\Lambda$  is defined by the semiempirical method and contains the parameters;<sup>3</sup> for the sake of compactness, we will write

$$\boldsymbol{\Theta} = \boldsymbol{\Lambda} \mathbf{P}$$

In order to derive the ddCOSMO contribution to the Fock matrix, we can exploit the chain rule

$$\mathbf{F} = \frac{\partial E^s}{\partial \mathbf{P}} = \frac{\partial E^s}{\partial \boldsymbol{\Theta}} \frac{\partial \boldsymbol{\Theta}}{\partial \mathbf{P}} = \boldsymbol{\Lambda}^\dagger \boldsymbol{\Xi},$$

where we introduced the “atomic Fock matrix”

$$[\Xi_i]_l^m = \frac{\partial E^s}{\partial [\Theta_i]_l^m}. \quad (7)$$

Therefore, we can compute the Fock matrix by assembling the derivatives of the solvation energy with respect to the atomic multipoles and the transform it in the AO basis through  $\boldsymbol{\Lambda}$ . By differentiating eq. 5:

$$[\Xi_i]_l^m = \sqrt{\frac{2l+1}{4\pi}} \delta_{l \leq 2} [\sigma_j]_l^m + \sum_j \sum_{l'm'} \sqrt{\frac{2l+1}{4\pi}} [\Theta_j]_{l'}^{m'} \frac{\partial [\sigma_j]_{l'}^{m'}}{\partial [\Theta_i]_l^m}$$

In order to compute the derivatives of  $\boldsymbol{\sigma}$  with respect to the atomic multipoles, we differ-

entiate the ddCOSMO linear system of equations  $\mathbf{L}\boldsymbol{\sigma} = \mathbf{g}$ . The ddCOSMO matrix is a block-sparse matrix: for each pair of spheres  $j, k$ :

$$[L_{jk}]_{ll'}^{mm'} = - \sum_{n=1}^{N_g} w_n Y_l^m(\mathbf{y}_n) W_n^{jk} \frac{4\pi}{2l'+1} (t_n^{jk})^{l'} Y_{l'}^{m'}(\mathbf{s}_n^{jk}).$$

Here,  $\{\mathbf{y}_n, w_n\}_{n=1}^{N_g}$  is the set of  $N_g$  points and weights of an appropriate Lebedev integration grid,  $W_n^{jk}$  is a weight depending on the overlap of the two spheres (it vanishes if the spheres do not intersect each other),  $t_n^{jk}$  is the distance from the center of sphere  $k$  to the  $n$ -th grid point on sphere  $j$  divided by the radius of sphere  $k$ ,  $\mathbf{s}_n^{jk}$  is the unit vector pointing from the center of sphere  $k$  to the same point and  $Y_l^m$  is a real spherical harmonic. Furthermore,

$$[g_i]_l^m = \sum_n w_n Y_l^m(\mathbf{y}_n) U_n^i \Phi_n^i,$$

where  $U_n^i$  is a weight introduced in order to regularize the problem and  $\Phi_n^i$  is the solute's electrostatic potential at the  $n$ -th grid point on the  $i$ -th sphere and is therefore the only quantity that depends explicitly on the atomic multipoles:

$$\frac{\partial[\sigma_j]_{l'}^{m'}}{[\Theta_i]_l^m} = \left([L_{ji}]_{l'l}^{m'm}\right)^{-1} \sum_n w_n Y_l^m(\mathbf{y}_n) U_n^i \frac{\partial\Phi_n^i}{\partial[\Theta_i]_l^m}$$

Now, let  $[\Psi_i]_m^l = \sqrt{\frac{4\pi}{2l+1}} [\Theta_i]_l^m$  and

$$\mathbf{s} = (\mathbf{L}^\dagger)^{-1} \boldsymbol{\Psi}$$

By putting everything together, eq. 8 is recovered:

$$[\Xi_i]_l^m = \frac{\partial E^s}{\partial[\Theta_i]_l^m} = \frac{1}{2} f(\varepsilon) \left[ \sqrt{\frac{4\pi}{2l+1}} [\sigma_i]_l^m \delta_{l \leq 2} + \sum_{j=1}^M \sum_{l'm'} [s_j]_{l'}^{m'} \frac{\partial[g_j]_{l'}^{m'}}{\partial[\Theta_i]_l^m} \right] \quad (8)$$

## References

- (1) Cancès, E.; Maday, Y.; Stamm, B. Domain decomposition for implicit solvation models. *J. Chem. Phys.* **2013**, *139*, 054111.
- (2) Lipparini, F.; Stamm, B.; Cancès, E.; Maday, Y.; Mennucci, B. Fast Domain Decomposition Algorithm for Continuum Solvation Models: Energy and First Derivatives. *J. Chem. Theory Comput.* **2013**, *9*, 3637–3648.
- (3) Khandogin, J.; Gregersen, B. A.; Thiel, W.; York, D. M. Smooth Solvation Method for d-Orbital Semiempirical Calculations of Biological Reactions. 1. Implementation. *J. Phys. Chem. B* **2005**, *109*, 9799–9809.