

# Chapter 8

## Systems in Solution

**Benedetta Mennucci, Jacopo Tomasi**

Dipartimento di Chimica e Chimica Industriale, Università di Pisa, via Risorgimento 35,  
56126 Pisa, Italy email: bene@dccci.unipi.it

**Maurizio Cossi**

Dipartimento di Chimica, Università Federico II, via Cintia, 80126 Napoli, Italy email:  
cossi@chemistry.unina.it<sup>1</sup>

### Introduction

The importance of getting a good understanding of surrounding media effects on chemical systems can hardly be overestimated. Applications range from condensed phase chemistry to biochemical reactions *in vitro* and to biological systems *in vivo*.

With the enormous progress achieved by computer technology, an increasing number of models and phenomenological approaches are being used to describe the effects of a given surrounding medium on the electronic properties of selected subsystem. A number of quantum chemical methods and programs, currently applied to calculate *in vacuo* systems, have been supplemented with a variety of solvation models.

Theoretical studies about the solvent effects on the properties and behavior of molecules are generally performed according to a large variety of methods; however, for our scopes, it is useful to classify them according to two basically different strategies. In the former, which we shall not treat at all, we can collect supermolecule calculations and computer

---

<sup>1</sup>L.A. Montero, L.A. Díaz and R. Bader (eds.), **Introduction to Advanced Topics of Computational Chemistry**, 185 - 217, 2003, ©2003 Editorial de la Universidad de La Habana, Havana.

simulations. Even if the philosophy of these two approaches is quite different, in both cases one looks for a detailed description of the disposition and structure of molecules composing the liquid system.

The latter strategy, often complementary to the former, collects methods in which a target subsystem, the “solute” (possibly supplemented by a few nearby solvent molecules) is described at the microscopic level, while a secondary subsystem (“the solvent”) is modeled as an infinite macroscopic continuum medium having suitable properties.

Several approaches can be classified in this second domain. Some among them belong to the category of semiempirical methods and some play a role in the study of complex solutes (e.g. molecules of biological interest). Other retain the formalism of the *in vacuo ab initio* molecular calculations, including in the Hamiltonian an explicit expression of the solute-solvent potential. In the present work we shall focus our attention almost exclusively on this last approach, also known as Effective Hamiltonian Method (EHM).

### a) Theoretical Aspects and Computational Methods

## 8.1 The Effective Hamiltonian and the Free Energy

The effective Hamiltonian method has been widely used to study solvent effects in chemical and biological systems. In this method the solvent  $S$  is represented by a homogeneous continuum medium which is polarized by the solute  $M$  placed in a cavity built in the bulk of the dielectric. The solute-solvent interactions are described in terms of a solvent reaction field.

In the quantum-mechanical (QM) version of the method the solute molecule is studied *ab initio* and the interactions with the solvent are taken into account through an interaction potential  $\hat{V}_R$  which acts as a perturbation on the solute Hamiltonian:

$$\hat{H}^0 \Psi^0 = E^0 \Psi^0 \quad \textit{in vacuo} \quad (8.1)$$

$$\left[ \hat{H}^0 + \hat{V}_R \right] \Psi = E \Psi \quad \textit{in solution} \quad (8.2)$$

where  $\hat{H}^0$  is the Hamiltonian of the solute *in vacuo* and  $\Psi^0$  and  $\Psi$  are the solute wave functions *in vacuo* and in solution, respectively.

In eq.(8.1) the Born-Oppenheimer approximation is employed. This implies the standard partition of the Hamiltonian into an electronic and a nuclear part, as well as the factorization of the wave function into an electronic and a nuclear component. In this approximation eqs. 8.1 and 8.2 refer to the electronic wave function, and the electronic operators  $\hat{H}^0$  and  $\hat{V}_R$  parametrically depend on the nuclear coordinates.

The definition of the interaction potential  $\hat{V}_R$  implies the knowledge of a thermally averaged distribution function of the solvent molecules,  $g_S$ . The details of this averaged distribution may change in the various realizations of the model, keeping fixed the continuity of the function itself. Actually, the basic continuum model involves a simplified form of the distribution function;  $g_s$  is reduced to describe a linear isotropic continuum, characterized by the static dielectric constant  $\epsilon$  of the bulk solvent, which, in turn, depends on temperature and pressure.

Since the perturbation operator  $\hat{V}_R$  in turn linearly depends on the solute wave function  $\Psi$ , the Schrödinger equation 8.2 is not linear. In these conditions it can be shown that the variational energy has to be calculated by minimizing the functional

$$\mathcal{G}(\Psi) = \left\langle \Psi \left| \hat{H}^0 + \frac{1}{2} \hat{V}_R \right| \Psi \right\rangle \quad (8.3)$$

and that the resulting quantity has the status of a free energy. Then the free energy of a molecule in solution can be written

$$G^{el} = \left\langle \Psi \left| \hat{H}^0 \right| \Psi \right\rangle + \frac{1}{2} \left\langle \Psi \left| \hat{V}_R \right| \Psi \right\rangle + V_{NN} + \frac{1}{2} U_{NN} \quad (8.4)$$

where  $V_{NN}$  is the usual nuclear repulsion energy, and  $U_{NN}$  is the interaction energy between the solute nuclei and the solvent reaction field generated by the solute nuclei themselves (the factor  $\frac{1}{2}$  ensures that this contribution is a free energy).

Usually, in  $\hat{V}_R$  one includes electrostatic solute solvent interactions only: other contributions, due to steric, dispersive and repulsive interactions can be added to obtain the free energy in solution

$$G = G^{el} + G^{ster} + G^{dis} + G^{rep} \quad (8.5)$$

In many computational approaches,  $G^{ster}$ ,  $G^{dis}$  and  $G^{rep}$  are calculated by classical algorithms, modifying the solute energy but not its wave function. However, recent developments allow one to include in the solute Hamiltonian some terms related to dispersion-repulsion interactions.

Before examining in detail the various contributions to the free energy and their computational expressions, in the next section we shall illustrate how the whole system is partitioned between solute and solvent.

## 8.2 The solute cavity and its tessellation

In all the EHM approaches, one defines an empty cavity in the dielectric medium in which the solute  $M$  resides. The shape and size of the cavity are critical factors in the elaboration

of a method. A cavity with a wrong shape introduces distortions in the description of the reaction field and of the related solvent effects.

The cavity shapes actually employed can be classified as follows:

1. regular shapes (i.e. spheres, ellipsoids, and cylinders);
2. molecular shapes given by the union of overlapping spheres;
3. molecular shapes obtained by exploiting the definition of specific molecular shape functions;
4. isodensity surfaces.

In the following we shall limit our analysis to the molecular shapes only, which are the most used in the methods we shall describe in the next sections. About the procedures corresponding to the item 4 of our classification, namely those giving cavities determined directly from solute isodensity surfaces, we limit to recall that in their most refined versions they may evaluate the isodensity surface in a self consistent way with the quantum mechanical calculation of the solute wave function [1].

Before going inside the technical details of the single procedures giving cavities of type 2) and 3), here represented by the two computational techniques called GEPOL [2] and DEFPOL [3] respectively, a further classification must be introduced: in general we can define three main kinds of surfaces:

1. the proper van der Waals surface ( $S_w$ ), which is the external surface resulting from a set of spheres centered on the atoms or group of atoms forming the solute;
2. the solvent accessible surface ( $S_a$ ), defined as the surface generated by the center of the solvent, considered as a rigid sphere, when it rolls on the van der Waals surface;
3. the solvent excluding surface ( $S_e$ ), which can be defined as the contact surface of a probe sphere (with radius equal to the molecular radius of the solvent molecules) rolling on the van der Waals surface.

It is easy to see that  $S_a$  can be equivalently defined as a  $S_w$  with radii enlarged by a quantity equal to the radius of the solvent, and that the last definition corresponds to the molecular cavity defined by Richards [4].

The choice of the type of surface to be used depends on the specific solute-solvent interactions which are taken into account in the calculation, as we shall see in the following.

### 8.2.1 GEPOL

GEPOL [2] describes van der Waals (Sa) cavities in terms of spheres centered on solute atoms, and it approximates solvent excluding surfaces (Se) by adding some additional spheres to Sa. Se's are created with a sequential algorithm: when two sphere of the Sa cavity are close enough to exclude the solvent from the space between them, one or two additional spheres, not centered on atoms, are added, and the procedure is repeated to a prefixed threshold, considering all the possible couples of spheres.

The surface of each sphere is then partitioned into triangular tesserae, corresponding to the projection of the faces of a suitable polyhedron inscribed in the sphere. This polyhedron may be defined in different ways, for example a geodesic partition based on pentakisdodecahedron giving origin to 60, 240, or 960 faces can be used (numerical tests have shown that the value of 60 is the best compromise between effectiveness and fastness, and this is the partition most used in the practice). Tesserae with all the vertices inside the cavity are discarded, while those whose surface is partially covered by some other spheres are replaced by suitable polygonal tesserae. In the most recent version of GEPOL [5] the area  $a_k$  of the polygonal tessera  $k$  is calculated by applying the Gauss-Bonnet theorem:

$$a_k = R_k^2 \left[ 2\pi + \sum_n^{N_k} \Phi_n \cos \vartheta_n - \sum_n^{N_k} \omega_n \right] \quad (8.6)$$

where  $R_k$  is the radius of the sphere which the tessera  $k$  belongs to,  $\Phi_n$  the angle subtending the edge  $n$  of the polygon (there are  $N_k$  edges),  $\vartheta_n$  the polar angle, and  $\omega_n$  the exterior angle of vertex  $n$ .

The most important characteristic of this formula is that it allows an analytical expression for the first derivatives of  $a_k$  with respect to all the relevant nuclear coordinates, including nucleus  $k$  but also the other nuclei defining the spheres which cut tessera  $k$  (actually, the derivatives also depend on the position and radius of the added spheres eventually cutting  $k$ ) [5]. The computation of this quantity, as we shall see in a following section, is a necessary step in the calculation of free energy derivatives with respect to nuclear coordinates as requested by standard geometry optimization techniques.

### 8.2.2 DEFPOL

This procedure [3], giving molecular cavities of type 3), starts from the definition, for the solute  $M$  at a given geometry, of an appropriate shape function whose value is 1 if calculated inside the molecular cavity and 0 otherwise. The molecular shape function is expressed as the union of hard-sphere shape functions  $f_{hs}(A)$  centered on the atoms of  $M$ :

$$f_{hs}(M) = \bigvee_{A \in M} f_{hs}(A) \quad (8.7)$$

where  $\bigvee$  is the logical operator "or" (inclusive) which runs over the atoms of the solute, and  $f_{hs}(A)$  the atomic shape function, related to the van der Waals radius, eventually

multiplied by a suitable numerical factor, of the atomic sphere and depending on the position of atom  $A$ . The molecular shape function given in eq.(8.7) is sufficient to define  $S_w$  or  $S_a$ , the only change needed is on the value of the scaling factor multiplying each radius, but not to define  $S_e$ . The latter has to be computed by supplementing  $f_{hs}(M)$  with a further shape function, called “obstruction shape function”,  $f_{obs}(M)$ :

$$f_{S_e}(M) = f_{obs}(M) \bigvee f_{hs}(M) \quad (8.8)$$

The analytical determination of  $f_{obs}(M)$  can be obtained by resorting to a cluster expansion, i.e. a sum of contributions due to pairs, triplets, etc., of atoms (it can be shown that the expansion can be limited to pairs and triplets). Having thus defined the suitable shape functions, the next step is to inscribe a regular polyhedron with planar triangular faces (regarding its choice, considerations given for GEPOL are valid also here) into a unit sphere centered on the center of mass of the molecule, and to adapt it with a sequence of deformations. The first deformation consists in the projection of the polyhedron on the main inertia ellipsoid of the molecule. The vertices of triangular faces are then shifted toward the molecular center of mass so as to touch the surface of the shape function. This shift is followed by the definition of an auxiliary point, placed at the intersection of the shape function surface with the axis perpendicular to the triangle and passing through its center. The sphere defined by these four points is then defined, and the planar triangle is replaced by its spherical counterpart. At the end of the whole process one obtains a tessellation of the cavity in terms of spherical triangles only (and not, as GEPOL, in spherical triangles supplemented by spherical polygons), but with both concave and convex curvature.

## 8.3 The contributions to the free energy

### 8.3.1 Steric contribution.

This term, whose contribution to  $G$  is more often called ‘cavitation free energy’,  $G^{cav}$ , corresponds to the work spent in building up a cavity of appropriate shape and volume in which the solute molecule is enclosed, with all the other solute-solvent interactions switched off: its effects on the system are only of energetic nature.

For the evaluation of  $G^{cav}$  several formulas are available, based on the shape and size of the solute and on different parameters of the solvent: surface tension [6], surface tension with microscopic corrections [7], isothermal compressibility [8], and geometrical data of the molecules [9, 10]. The first three formulas here mentioned are of empirical nature and follow almost the same philosophy of the continuum dielectric, neglecting the discrete nature of the solvent molecules but making use of experimental bulk parameters. The last formulation, on the contrary, derives from a theory based on a discrete model of fluids (the Scaled Particle Theory, SPT), even if the final expression of  $G^{cav}$  depends again on bulk solvent parameters only.

The common characteristic of all these different formulations is that their formal def-

initiation has been derived only for spherical, or at most ellipsoid, cavities. Actually, the largest part of the computational codes for calculations in solution makes use of more complex molecular cavities, often formed by many interlocking spheres; in these cases, the analytic formulas derived in references quoted above have to be rewritten with some approximations.

The procedure reported in ref. [9, 10] is that commonly used in solvation programs, then we shall explicitly report this expression only:

$$G^{cav} = \sum_i^{spheres} \frac{A_i}{4\pi R_i^2} G^{cav}(R_i) \quad (8.9)$$

where we have supposed that the molecular cavity is formed by interlocking spheres, each having a radius  $R_i$  and contributing with a weight related to the portion of its surface which is exposed to the solvent ( $A_i$ ); the spherical limit value,  $G^{cav}(R_i)$ , is written as an expansion in powers of  $R_{MS}$ , i.e. the radius of the sphere which excludes the centers of the solvent molecules (sum of the solute and solvent radii,  $R_{MS} = R_M + R_S$ ):  $G^{cav} = \sum_{x=0}^3 K_x R_{MS}^x$ , being the coefficient  $K_x$  defined in terms of solvent molecular radius and numeral density, and solution pressure and temperature.

The cavitation energy is calculated using a van der Waals surface (Sw).

### 8.3.2 Repulsion and dispersion contributions.

What we are calling repulsion contribution could be considered as another part of the steric contribution described above. However, it is treated together with the dispersion contribution, although of different physical origin, as their analysis and the consequent classification follow almost the same path; moreover, also in the classical literature on molecular interactions, they are often indicated with a single name, the van der Waals contributions.

The modeling of dispersion and repulsion interactions in solution may be based either on a discrete molecular description of the liquid or on a continuum dielectric model. Let us analyze the two approaches separately.

#### The discrete approach

The discrete approach is generally based on an operator  $\hat{V}_{dis-rep}$  written in terms of pair potentials related to atoms or groups of atoms of the solvent  $S$  (here indicated with  $s$ )

and the solute  $M$  (here indicated with  $m$ ):

$$\hat{V}_{dis-rep} = \sum_{m \in M} \sum_{s \in S} V_{ms}(r_{ms}) \iff V_{ms}(r_{ms}) = \sum_n \frac{d_{ms}^{(n)}}{r_{ms}^n} \quad (8.10)$$

the dispersion ( $n=6,8,10$ ) and the repulsion ( $n=12$ ) coefficients are taken from the literature. Often only the term with  $n = 6$  is retained for the dispersion and an alternative exponential expression, more related to the physical interpretation of the interaction, is used for the repulsion term:  $c_{ms} \exp(-\gamma_{ms} r_{ms})$ .

As for the previous steric term this potential is completely independent from solute charge distribution, hence its effect is only on the system free energy. This contribution to the free energy depends on the solvent atoms distribution function outside the cavity: in the isotropic approximation, this distribution function is constant and equal to the solvent numerical density  $\rho_{solv}$ , so that one can write [11]

$$G^{dis(rep)} = \rho_{solv} \sum_{s \in S} \sum_{m \in M} \left( \sum_i^{(tesserae)} a_i V_{ism}^{dis(rep)} \vec{r}_{mi} \cdot \hat{n}_i \right) \quad (8.11)$$

where the sum on  $i$  runs over the tesserae on the cavity surface,  $\vec{r}_{mi}$  is the distance vector between tessera  $i$  and solute atom  $m$  and  $\hat{n}_i$  is the unit vector normal to the cavity surface on tessera  $i$ . The terms  $V_{ism}^{dis(rep)}$  are

$$V_{ism}^{dis} = -\frac{1}{3} \frac{d_{ms}}{r_{mi}^6} \quad (8.12)$$

$$V_{ism}^{rep} = c_{ms} \exp(-\gamma_{ms} r_{mi}) \left[ \frac{1}{\gamma_{ms} r_{mi}} + \frac{2}{(\gamma_{ms} r_{mi})^2} + \frac{2}{(\gamma_{ms} r_{mi})^3} \right] \quad (8.13)$$

Since the semi-empirical parameters entering the operator  $\hat{V}_{dis-rep}$  are defined in terms of distances between solute and solvent nuclei,[12] the term  $G^{dis}$  and  $G^{rep}$  is best calculated using a solvent accessible surface: in fact, Sa determines the closest approach of solvent nuclei to the solute.

### The continuum approach.

There exist some models which describe dispersive forces between molecules by quantum mechanical equations: we don't review them here, limiting ourselves to cite a recent extension [13], which permits the calculation of  $G^{dis}$  in the context of continuum solvation models. Such approach makes use of a suitable operator to be added to  $\hat{H}^0 + \hat{V}_R$ : then the dispersion solute-solvent interaction influences not only the solute energy, but also its electronic distribution.

In the same paper [13], a quantum mechanical approach for the calculation of the repulsion energy is also given, in terms of a simple operator proportional to the fraction



of solute electronic charge lying outside the cavity. Again, this description allows one to compute the effects of the repulsion interaction both on the solute wave function and on its energy.

### 8.3.3 Electrostatic contribution

Let us now analyze the last and most important contribution to the solute-solvent interactions, namely  $G^{el}$ . In this case a deeper attention must be paid to the macroscopic nature of the continuum medium representing the solvent. Effective Hamiltonian methods have been derived to describe different kinds of solvents:

1. homogeneous isotropic dielectrics, characterized by a constant scalar permittivity,  $\epsilon$ ;
2. homogeneous isotropic dielectrics in the presence of other charged solutes (ionic solutions);
3. homogeneous anisotropic dielectrics, characterized by a constant tensorial permittivity,  $\epsilon$ ;
4. inhomogeneous dielectrics, characterized by a position dependent permittivity,  $\epsilon(\vec{r})$ .

Let us consider the most simple system, namely a solute  $M$  with nuclear and electronic charge distribution  $\rho_M$  in a cavity  $C$  surrounded by an infinite homogeneous isotropic dielectric with permittivity  $\epsilon$ : the basic relation to be considered is the Poisson equation with the related boundary conditions, namely

$$\left\{ \begin{array}{ll} -\Delta V = 4\pi\rho_M & \text{in } C \\ -\epsilon\Delta V = 0 & \text{outside } C \\ V_i = V_e & \text{on } \Sigma \\ \left(\frac{\partial V}{\partial n}\right)_i = \epsilon\left(\frac{\partial V}{\partial n}\right)_e & \text{on } \Sigma \end{array} \right. \quad (8.14)$$

where  $V$  is the total electrostatic potential,  $\Sigma$  is the cavity surface, the subscripts  $e$  and  $i$  indicate regions outside and inside the cavity, respectively, and  $n$  is the (outward) unit vector perpendicular to the cavity.

The three most widely disseminated approaches used to solve this problem are

1. the apparent surface charge (ASC) methods: PCM [14, 15, 16], COSMO [17, 18], IEF [19, 20];
2. the multipole expansion (MPE) methods: SCRF [21, 22];
3. the generalized Born approximation (GBA) methods: AMSOL [23].

In the list above each approach is associated with one or more specific computational procedures, indicated with their acronyms, which represent the most successful applications of the class itself. Let us sketch the AMSOL and SCRF procedures briefly, before reviewing the ASC methods more accurately.

### GBA methods (AMSOL).

These methods can be defined as the generalization of the Born model [24]] formulated for the simple system of a point charge  $q$  placed at the center of a void spherical cavity with radius  $a$  (or equivalently a conducting sphere with net charge  $q$  and radius  $a$ ). The extension of the Born formula giving the electrostatic contribution to solvation free energy to the by far more complex system of an assembly of point charges in a general cavity can be written as follows:

$$G_P(Born) = -\frac{\epsilon - 1}{2\epsilon} \frac{q^2}{a} \Rightarrow G_P(GBA) = -\frac{\epsilon - 1}{2\epsilon} \sum_{k,k'} q_k q_{k'} \gamma_{kk'} \quad (8.15)$$

where  $k, k'$  label the atomic centers and  $\gamma_{kk'}$  are Coulomb-type integrals.

The same formalism, with the required changes, has been adopted by Cramer and Truhlar [25] in a series of methods, now included in the AMSOL code, which belong to the same class of effective Hamiltonian exploiting continuum description of the solvent we have defined at the very beginning.

In these methods, called  $SMx$  where  $x$  denotes the parametrization of the specific solvation method, the partial atomic charges  $q_k$  to be used in the application of eq.(8.15) are calculated either from a semiempirical wave function using a zero-overlap Mulliken population analysis or, more recently, from a class of charge models which map the Mulliken derived charges with others reproducing more accurately experimental gas-phase dipole moments.

In the most recent versions of  $SMx$  models, the solvation free energy is calculated as the sum of two terms:

$$\Delta G_{sol} = \Delta G_{ENP} + G_{CDS} \quad (8.16)$$

where the first term  $\Delta G_{ENP}$  contain, on turn, two components:  $\Delta E_{ENP}$ , representing the change in the internal electronic kinetic and electronic-nuclear coulombic energies of the solute upon relaxation in solution, and  $G_P$  which is the free energy of the electric polarization as derived from the generalized Born approximation. The second term of eq.(8.16),  $G_{CDS}$ , accounts both for the free energy of forming a cavity in the solvent to make room for the solute, and for the changes in dispersion interactions and solvent structure that always accompany the solvation process.

Here we don't consider the  $CDS$  term, which is similar to the  $G^{cav}$  term seen above, and focus on the electrostatic term  $G_{ENP}$ .

The whole method is here based on a self-consistent reaction field formalism, in which the general QM equation to be defined is the following:

$$\mathbf{F} = \frac{\partial G}{\partial \mathbf{P}} \quad (8.17)$$

where  $\mathbf{F}$  is the Fock matrix,  $G$  the already introduced free energy functional, and  $\mathbf{P}$  the solute density matrix. For a restricted Hartree-Fock calculation in which core electrons are not treated explicitly, the free energy functional becomes:

$$G = \frac{1}{2} \sum_{\mu, \nu} P_{\mu\nu} (h_{\mu\nu} + F_{\mu\nu}^{(0)}) + \sum_{i < j} \frac{Z_i Z_j}{R_{ij}} - \frac{\epsilon - 1}{2\epsilon} \sum_{k, k'} q_k q_{k'} \gamma_{kk'} \quad (8.18)$$

where  $\mu$  and  $\nu$  run over the set of valence atomic orbital functions, and  $Z_i$  is the nuclear charge of atom  $i$  minus its number of core electrons. Here  $\mathbf{P}$  is the relaxed (with respect to solvation) density matrix,  $\mathbf{F}^{(0)}$  the Fock matrix as defined *in vacuo*, but formed using the relaxed density matrix, and the final term accounts for electric polarization using the generalized Born formalism. Note that the first term still represents the electronic energy of the solute, including electron-nuclear attraction exactly as *in vacuo*, although it will generally be higher in energy than it was the corresponding gas-phase term since the density matrix has been changed from the gas-phase optimum.

In order to arrive at the relaxed density matrix, one must solve for the orbitals with a proper Fock matrix as defined by eq.(8.17). Doing so requires to take the partial derivative of eq.(8.18) with respect to the density matrix. This yields:

$$F_{\mu\nu}^{(1)} = F_{\mu\nu}^{(0)} - \frac{\epsilon - 1}{\epsilon} \sum_{k, k'} q_k \left( \frac{\partial q_{k'}}{\partial P_{\mu\nu}} \right) \gamma_{kk'} - \frac{\epsilon - 1}{2\epsilon} \sum_{k, k'} q_k q_{k'} \left( \frac{\partial \gamma_{kk'}}{\partial P_{\mu\nu}} \right) \quad (8.19)$$

Once more, it is worth noting the differences between the operators, here expressed in their matrix form, used in the definition of the QM problem and those exploited in the evaluation of the free energy. In the present formalism, we have that matrix  $\mathbf{F}^{(1)}$  is required for solution of the SCF equations, but  $\mathbf{F}^{(0)}$  is used in the calculation of  $G_{ENP}$  according to eq.(8.18).

As a concluding comment, we emphasize, even if it should be already clear from the derivation above, that by explicitly including the polarization effects in the Fock operator, the resulting density matrix and converged orbitals are determined self-consistently in the presence of the solvent. Moreover AMSOL methods are fast, usable in different solvents, and provided with many options, among which geometry optimization. On the other hand, one must remember that these methods are strongly dependent on the type of parametrization used and, more important, that only semiempirical quantum procedures can be exploited, namely the Austin Model 1 (AM1), or the Parametrized Model 3 (PM3).

### MPE methods (SCRf).

This class of methods will be here described almost exclusively in terms of a specific model known as SCRf method [21, 22], which is without any doubt the most complete MPE

method.

Like GBA is defined as the generalization of the Born model, the MPE approach can be seen as an extension of Kirkwood's model [26]. Here, in fact, the charge distribution of the solute placed in a cavity surrounded by a continuum dielectric is expanded in a series of multipoles: in this framework, if with  $M_m^l$  we indicate the  $m$  component of the multipole of order  $l$  written in the spherical tensor formalism, the electrostatic contribution to the free energy variation becomes:

$$\Delta G_{el} = -\frac{1}{2} \sum_{l=0}^{\infty} \sum_{m=-l}^l R_l^m M_l^m \quad (8.20)$$

in which  $R_l^m$  is a component of the reaction field, i.e. the corresponding derivatives of order  $l$  of the electrostatic potential created by the continuum polarized under the influence of the solute.

Within the linear response approximation the various components of the reaction field can be written in the form:

$$R_l^m = \sum_{l',m'} f_{ll'}^{m,m'} M_{l'}^{m'} \quad (8.21)$$

in which the coefficients  $f_{ll'}^{m,m'}$ , called the reaction field factors, only depend on the shape of the cavity and of the dielectric constant of the solvent.

In original Kirkwood's treatment, using a spherical cavity, these factors have a simple analytical definition: they are independent from  $m$  and are nonzero only when  $l = l'$ . The analytical treatment is possible also for a spheroid or an ellipsoid. In the general case, the reaction field factors have to be computed numerically.

The formalism given above for a one-center expansion can be, and actually has been, straightforwardly generalized to a distributed multipole expansion of  $N$  centers; this development, as expected, leads to much more rapidly convergent calculations of electrostatic solvation free energies. However, for simplicity's sake, in the following we shall continue to limit the exposition to the one-center expansion formalism.

Once again, in order to pass to the quantum calculation, here limited to the description of the solute electronic wave function as a single antisymmetrized product of one-electron molecular spin-orbitals expanded over a finite set of atomic orbitals  $\{\chi_\mu\}$ , we have to minimize the free energy obtaining the modified Fock matrix whose  $(\mu\nu)$  element is written:

$$F_{\mu\nu} = F_{\mu\nu}^0 + \sum_{l,m} \sum_{l',m'} M_l^m f_{ll'}^{m,m'} \langle \mu | S_{l'}^{m'} | \nu \rangle \quad (8.22)$$

in which  $F_{\mu\nu}^0$  is the corresponding matrix element for the isolated molecule and  $S_{l'}^{m'}$  a solid spherical harmonic which defines the corresponding multipole tensor element. The perturbation term of eq.(8.22), because of its nonlinearity, depends on the density matrix  $\mathbf{P}$ , as it can be easily shown by writing the explicit definition of each  $M_l^m$  term:

$$M_l^m = \sum_{\lambda} \sum_{\eta} P_{\lambda\eta} \langle \lambda | S_l^m | \eta \rangle \quad (8.23)$$

Nevertheless, this dependence does not introduce any special difficulty since the same characteristic occurs in  $F_{\mu\nu}^0$ . The iterative scheme in the SCF computation is adapted to this situation.

This methodology, originally developed for the evaluation of the electrostatic free energy only and limited to the simple case of a spherical cavity, more recently has been generalized to cavities of arbitrary shape, and to more complex calculations. It is worth mentioning that in a recent version of the code it has been introduced the calculation of analytical derivatives of the free energy [27], and hence the possibility of geometry optimization procedures, and that the same formalism of the reaction field factors allows a more comprehensive approach in which the dispersion term is added to the electrostatic term in the derivation of the Hartree-Fock equation [28].

### ASC methods (PCM, COSMO, IEF).

In order to derive the working quantities of each of the three methods chosen as representatives of this class, namely PCM, COSMO and IEF, it is worth going back to system 8.14. In fact, all these methods solve the system and get the related potential  $V$  by exploiting an integral equation formalism.

Actually, the COSMO model is a little different, since the screening effects in the dielectric are replaced by the screening effects in a conductor. In other words, the COSMO method is a solution of the Poisson equation designed for the case of very high  $\epsilon$ , and it takes advantage of the analytic solution for the limit case of a conductor ( $\epsilon = \infty$ ), for which the boundary condition reduces to  $V = 0$  on the surface. Anyway, apart from these differences in the theoretical background, COSMO can be described exactly in the same way as the other ASC methods, and, in fact, in the following we shall present a single derivation for all the methods.

The system (8.14) can be solved by a quite standard approach based on Green functions, or by a procedure very recently implemented using operator functions derived from the theory of integral equations. The former method is applied by PCM and COSMO, while IEF exploit the latter approach. In both cases, the differential equations of system 8.14 are transformed into integral equations on the surface  $\Sigma$ , that can be easily solved by standard numerical methods.

In this framework, the solution of system (8.14) is given by the sum of two electrostatic potentials, one produced by  $\rho_M$  and the other due to a surface charge distribution  $\sigma$ , placed on the interface, arising from the polarization of the dielectric medium:

$$V(x) = V_M(x) + V_{\sigma}(x) = \int_{R^3} \frac{\rho_M(y)}{|x-y|} dy + \int_{\Sigma} \frac{\sigma(s)}{|x-s|} ds \quad (8.24)$$

where the integral on the first term is taken over the whole space. The surface charge has to fulfill another condition derived from the Gauss theorem:

$$Q_{\sigma}(\text{theo}) = \int_{\Sigma} \sigma(s) ds = -\frac{\epsilon - 1}{\epsilon} Q_M \quad (8.25)$$

where  $Q_M$  is the total net charge of the solute  $M$ . The condition expressed by eq.(8.25) is clearly a great advantage of the ASC formulation with respect to all the other methods we have described in the previous subsections, in fact, it gives an easy checking test on the quality of the model employed.

In ASC approaches exploits the surface  $\Sigma$  is partitioned into tesserae of known area  $a_k$ , on which a constant charge density is assumed. In this framework, easily linkable to the analogous techniques used in the fields of physics and engineering and known as Boundary Element Methods (BEM), the integral form of  $V_{\sigma}(x)$  in eq.(8.24) is reduced to a finite sum running over the point charges representing the surface charge:

$$V_{\sigma}(x) = \sum_k^K \frac{q_k(s_k)}{|x - s_k|} \iff q_k(s_k) = \sigma(s_k) a_k \quad (8.26)$$

where vector  $s_k$  indicates the representative point of each tessera  $k$ , i. e. the point where  $\sigma$  is evaluated.

Let us consider the three ASC methods in detail, starting from PCM, which for its longer history can be taken as a reference for the definition of quantities used in the other methods, too.

**PCM** In this method, developed in its first form in 1981 [14], but then almost completely redefined in 1995 [16], the apparent surface charge is expressed by the following classical electrostatic relation:

$$\sigma(s) = -\left(\frac{\epsilon - 1}{4\pi\epsilon}\right) \vec{E}(s) \cdot \hat{n}(s) \quad (8.27)$$

where  $\vec{E}$  is the total electric field, i.e. the sum of the contributions due to the solute charge distribution and to the ASC itself, respectively, and  $\hat{n}$  is the outward normal unit vector to the cavity at position  $s$ .

In the computational practice  $\sigma$  is considered constant over each tessera, so that a set of apparent point charges  $q_i = a_i \sigma(s_i)$  are distributed on the cavity surface. Usually, the solute charge distribution,  $\rho_M$ , is partitioned into nuclear and electronic contributions, and the set of linear equations defining the apparent charges, below collected in the column matrix  $\mathbf{q}$  is given by:

$$\mathbf{D}\mathbf{q} = -(\mathbf{E}_M^e + \mathbf{E}_M^N) \quad (8.28)$$

where  $\mathbf{D}$  is a square matrix, with size equal to the number of surface tesserae, with elements

$$\begin{aligned} D_{ii} &= \frac{4\pi}{a_i} [\epsilon/(\epsilon - 1) - (1 - \xi_i)/2] \\ D_{ij} &= [(s_i - s_j) \cdot n(s_i)/|s_i - s_j|^3] \end{aligned} \quad (8.29)$$

The diagonal term of  $\mathbf{D}$ , which collects the contribution of the reaction field induced by the charge placed on tessera  $i$  on itself, is derived by the Gauss formula for an infinite charged plane with a correction term  $\xi_i$  accounting for the curvature of the convex tessera. The vectors  $\mathbf{E}_M^e$  and  $\mathbf{E}_M^N$  collect the normal components of the electric field coming from solute electrons and nuclei, respectively (the subscript  $M$  indicates that we are considering the field generated by the solute only).

The column matrices on the right hand side of eq. 8.28 collect the values of the normal electric field due to solute nuclei and solute electrons, respectively.

Also apparent charges can be partitioned into two components defined as electron- and nuclei-induced, respectively:

$$\mathbf{q} = \mathbf{q}^e + \mathbf{q}^N = -\mathbf{D}^{-1} \left( \mathbf{E}_M^e + \mathbf{E}_M^N \right) \quad (8.30)$$

Thus the solute-solvent interaction energy,  $U_{MS}$  may be divided into four different contributions:

$$U_{MS} = \int \int \frac{\rho(x)\sigma(s)}{|x-s|} dx ds = U_{ee} + U_{eN} + U_{Ne} + U_{NN} \quad (8.31)$$

where the first subscript refers to the component of the solute charge and the second to that of the ASC.

Limiting for simplicity to closed shell Hartree-Fock calculations, the first three contributions to the interaction energy of eq.(8.31) (the last term  $U_{NN}$  is a constant dependent only on the positions and the charges of solute nuclei) can be expressed in terms of three matrices,  $\mathbf{X}(\mathbf{P})$ ,  $\mathbf{j}$ , and  $\mathbf{y}$ , collecting solute-solvent interaction terms, so that:

$$\begin{cases} U_{ee} = tr\mathbf{P}\mathbf{X}(\mathbf{P}) \\ U_{eN} = tr\mathbf{P}\mathbf{j} \\ U_{Ne} = tr\mathbf{P}\mathbf{y} \end{cases} \quad (8.32)$$

where  $\mathbf{P}$  is the solute one electron density matrix on a finite atomic basis set  $\{\chi\}$ . In particular,  $\mathbf{j}$  is the matrix collecting the interactions between each solute electronic elementary charge distribution  $\chi_\mu\chi_\nu$  and the nuclei-induced apparent charges  $\mathbf{q}^N$ ,  $\mathbf{y}$  the matrix related to the interaction between solute nuclear charges and the electron-induced apparent charges  $\mathbf{q}^e$ , and  $\mathbf{X}(\mathbf{P})$  that defining the interactions between solute electrons and the apparent charges  $\mathbf{q}^e$  [16].

The expression for the corresponding electrostatic component of the free energy can be easily written as:

$$\begin{aligned} G^{el} &= tr\mathbf{P} \left[ \mathbf{h} + \frac{1}{2}(\mathbf{j} + \mathbf{y}) \right] + \frac{1}{2}tr\mathbf{P} [\mathbf{G}(\mathbf{P}) + \mathbf{X}(\mathbf{P})] + \left[ \frac{1}{2}U_{NN} + V_{NN} \right] \\ &= tr\mathbf{P}\mathbf{h}' + \frac{1}{2}tr\mathbf{P}\mathbf{G}'(\mathbf{P}) + V'_{NN} \end{aligned} \quad (8.33)$$

where  $\mathbf{h}$ ,  $\mathbf{G}(\mathbf{P})$  and  $V_{NN}$  are the elements used in standard calculations *in vacuo*, collecting one-, two-electron integrals, and nuclear repulsion, respectively. Here, the prime on the matrices in the right hand expression, which is formally identical to that of  $E^0$  for the

system *in vacuo*, stresses that the corresponding term takes into account the contribution due to the solvent.

By combining the stationary condition  $\delta G_{el} = 0$  for an arbitrary variation of the molecular orbital coefficients  $\mathbf{C}$ , with the auxiliary conditions of orthonormality, one easily arrives to the following HF-like equation:

$$\mathbf{F}'\mathbf{C} = [\mathbf{h}' + \mathbf{G}'(\mathbf{P})] \mathbf{C} = \mathbf{S}\mathbf{C}\epsilon \quad (8.34)$$

where, once again, the prime on the Fock matrix indicates the presence of terms due to the solvent. Here  $\mathbf{S}$ , and  $\epsilon$  represent the standard overlap and one-electron orbital matrices, exactly as *in vacuo*.

Eq.(8.34) can be solved with the same iterative procedure of the problem *in vacuo*; the only difference introduced by the presence of the continuum dielectric is that, at each SCF cycle, one has to simultaneously solve the standard quantum mechanical problem and the additional electrostatic problem of the evaluation of the interaction matrices, and hence of the apparent charges. The latter are obtained from eq.(8.30) through a self-consistent technique which has to be nested to that determining the solute wave function; as a consequence, in each cycle solute and solvent distribution charges are mutually equilibrated.

However, we have to stress that in the computational practice many numerical problems can disturb the calculation; the most important is that Gauss' condition on the total apparent charge (see eq.8.25) is generally not fulfilled. There are two main reasons for this important discrepancy: a numerical one, due to the discretization of the apparent surface charge distribution into a finite set of point charges, and a physical one, related to the solute electronic charge spreading out of the cavity. The numerical error affects both  $\mathbf{q}^N$  and  $\mathbf{q}^e$  (more or less at the same extent), while the physical error is referred to  $\mathbf{q}^e$  only.

In general, one needs a process of renormalization of the apparent charges such as to bring the corrected  $\bar{q}_k^x$  to satisfy the condition:

$$\sum_k^K \bar{q}_k^x = -\frac{\epsilon - 1}{\epsilon} Q_M^x \quad (8.35)$$

where  $x$  stands for  $N$  or  $e$ . Since the first paper on PCM, this problem has been pointed out and many techniques of renormalization, also called compensation procedures, have been given [14, 16].

The simplest compensation corrects the two sets of charges through two separate factors  $f^e$  and  $f^N$ , constant in each set, such that  $\bar{q}_k^x = f^x q_k^x$  for all  $k$ .

In a more accurate treatment, the  $\mathbf{q}^N$ 's are corrected by the same  $f^N$  as in the preceding method, while the  $\mathbf{q}^e$ 's are first multiplied by  $f^N$  (in the hypothesis that the numerical error is equal for the two sets of charges), and then the remaining error (mainly due to escaped electrons) is corrected by an additional apparent charge. Such additional charge is proportional to the solute electron density on the tesseræ: in such a way, the corrections



are greater in those parts of the cavity where the flux of electronic tails through the surface is more pronounced.

A third, more refined compensation procedure has been recently proposed [29]: the factor  $f^N$  is used as in the previous method, then the escaped electrons are compensated by means of an additional "effective" apparent charge,  $\sigma_{eff}$ . In this case  $\sigma_{eff}$  is treated exactly as the usual polarization charge, i. e. it is partitioned into finite point charges and used to modify the Hamiltonian.

More details about compensation can be found in ref. [30].

**COSMO** This method, originally proposed [17, 18] in a form quite different from PCM, has been recently recast [31] in order to exploit the same features elaborated for PCM.

As already said, COSMO is based on a screening conductor theory. Screening in conductors can be handled more easily than in dielectric media; in our framework, i. e. a solute  $M$  placed in a cavity inside the conductor, the boundary condition to be fulfilled is that the total electrostatic potential  $V$  cancels out on the cavity surface

$$V(s) = V_M(s) + V_\sigma(s) = V_M(s) + \sum_k^K \frac{q_k(s_k)}{|s - s_k|} = 0 \quad \forall s \in \Sigma \quad (8.36)$$

where we have exploited a partition of the cavity surface into tesserae which allows to discretize the surface charge distribution  $\sigma$  into a set of  $K$  point charges.

Following the strategy already used to get eq. (8.28) in the PCM method, from eq.(8.36) we can write the basic system giving the COSMO apparent charges in the following matrix formalism:

$$\mathbf{BQ} = -\mathbf{V}_M \quad (8.37)$$

where the column matrix  $\mathbf{V}_M$  contains the solute electrostatic potential on each tessera, and the elements of the square matrix  $\mathbf{B}$  are:

$$\begin{aligned} B_{ii} &= 1.07a_i\sqrt{4\pi/a_i} \\ B_{ij} &= a_j/|s_i - s_j| \end{aligned} \quad (8.38)$$

with quantities having the same meaning seen above. The numerical factor 1.07 is derived from the simple case of a homogeneously charged sphere [17].

In equation 8.37 we used  $\mathbf{Q}$  to highlight that this apparent charges are obtained for a conductor, i. e. for  $\epsilon = \infty$ ; if the COSMO model is used to simulate a solvent with dielectric constant  $\epsilon$ , these apparent charges have to be scaled so that their sum obeys the Gauss law (see eq.8.25): this can be effectively accomplished by multiplying each charge by the factor  $(\epsilon - 1)/\epsilon$ . By exploiting the already defined partition into electrons and

nuclei-induced apparent charges, we can write:

$$\mathbf{q} = \mathbf{q}^e + \mathbf{q}^N = -\frac{\epsilon - 1}{\epsilon} \mathbf{B}^{-1}(\mathbf{V}_M^e + \mathbf{V}_M^N) \quad (8.39)$$

In a finite basis matrix formulation, the expression of the free energy functional from which one derives the quantum mechanical system to be solved is identical to that presented for the PCM model. We do not repeat here the whole derivation, but we only stress that also here the solvent reaction field is included directly in the SCF procedure, and hence the solute electron density distribution and the solvent reaction field are converged simultaneously.

Also COSMO is affected by the numerical and escaped electrons errors illustrated for PCM, however, it can be shown that the conductor approach is less sensitive to the problem of the escaped charge than the dielectric one. As a matter of fact, the error on the electrons-induced set of apparent charges generally reduces of an order of magnitude with respect to that found in PCM.

The same compensation procedures defined for PCM, even if specifically designed for a dielectric-like model, could be exploited also in the COSMO algorithm; moreover, a further technique explicitly thought for the conductor model has been recently presented [32]. Anyway, a large number of test calculations have shown that COSMO results are effectively corrected both for numerical and for physical errors by simply multiplying by constant factors (first procedure illustrated for PCM), even for charged solutes.

As a final note we add that COSMO method is rightly defined, and indeed it well reproduces solute energies and properties, only for highly polar solvents like water, for which the difference between the two values of  $(1-1/\epsilon)$  obtained with  $\epsilon = 78.5$  and  $\epsilon = \infty$  is only 1.3%. Nevertheless, sometimes COSMO is used also for the treatment of nonpolar solvents with  $\epsilon \approx 2$ , but in these cases it needs approximations which are physically by far less founded than those defined for the dielectric model, and not always easily acceptable.

**IEF** This method of very recent implementation [19, 20] needs few further comments on the problem of the operator formalism used to solve Poisson equation (8.14). In the first part of this section we underlined the fact that, while both PCM and COSMO exploit a mathematical approach based on Green functions, this new method makes use of other operators, well-known in the theory of integral equations [33] but not so diffused in the theoretical treatment of chemical problems. We do not report here their formal derivations, as this would be too long and almost out the scope of this chapter; what we want to stress is that this new formalism leads to two very important consequences. From one hand, it allows to treat in a single approach dielectrics of very different nature (i.e. standard isotropic liquids, anisotropic media like liquid crystals and solid matrices, and ionic solutions), and, on the other hand, it includes the two previous methods as particular cases: namely, the PCM can be derived from the standard isotropic description, and COSMO in the limit of an infinite dielectric constant.

Starting from suitable operators, and omitting all the complex derivation presented in refs. [19, 20], one obtains that the polarization point charges are determined by the matrix equation

$$\mathbf{C}\mathbf{q} = -\mathbf{g} \tag{8.40}$$

$\mathbf{q}$  is the vector of the polarization charges on tesserae,  $\mathbf{C}$  is a matrix defined as

$$\mathbf{C} = (\mathbf{I}/2 - \mathbf{D}_e) \mathbf{S}_i \mathbf{A}^{-1} + \mathbf{S}_e (\mathbf{I}/2 + \mathbf{D}_i^\dagger) \mathbf{A}^{-1} \tag{8.41}$$

where  $\mathbf{I}$  is the unit matrix,  $\mathbf{A}$  is a diagonal matrix collecting tesserae areas and  $\mathbf{D}_e$ ,  $\mathbf{D}_i$ ,  $\mathbf{S}_e$  and  $\mathbf{S}_i$  depend on the position of tesserae, on the solvent dielectric constant and, in the case of ionic solutions, on the ionic strength (their explicit expressions are given in ref. [20]). Lastly, the elements of the  $\mathbf{g}$  vector are combinations of the electrostatic potential ( $\mathbf{V}_M$ ) and of the normal component of the electric field ( $\mathbf{E}_M$ ) exerted by the solute on each tessera (see ref. [20] for more details):

$$\mathbf{g} = (\mathbf{A}/2 - \mathbf{D}_e) \mathbf{V}_M - \mathbf{S}_e \mathbf{E}_M \tag{8.42}$$

As usual, we can partition the apparent charges:

$$\mathbf{q} = \mathbf{q}^e + \mathbf{q}^N = -\mathbf{C}^{-1}(\mathbf{g}^e + \mathbf{g}^N) \tag{8.43}$$

Once again, the strategy and the tools used to derive the quantum mechanical problem are completely equivalent to that described for PCM and COSMO. In these terms, it is possible to perform an iterative calculation of the IEF formally equal to the standard SCF procedure *in vacuo*, leading to self-consistent polarization charges and solute wave function.

It is worth noting that IEF, exploiting both electrostatic potential and electric field, has a hybrid nature and it is more similar to PCM for some aspects and more similar to COSMO for others.

In particular, regarding the important problem of the escaped charge effects, the behavior shown by IEF is almost identical to that already seen in COSMO. Numerically, this means that the errors on IEF electrons-induced apparent charges are much smaller than the PCM ones and, at least for standard calculations on neutral solutes, they can be safely neglected without affecting the final results in a quantitative extent.

### 8.4 Free energy derivatives

Besides the “direct” effects (i. e. those related to the polarization of the solute wave function at the geometry optimized in *vacuo*) described in the previous sections, very often

one cannot neglect the geometry relaxation induced by the solvent (sometimes referred to as “indirect” solvent effects).

In the study of chemical reactions in vacuo it is compulsory to locate the critical points (minima and saddle points) on the potential energy hypersurface, and the reaction mechanisms can be characterized by following a suitable reaction coordinate. The analytical calculation of the energy gradients with respect to the nuclear coordinates and of the Hessian matrix are necessary to find and characterize the critical points and to define the intrinsic reaction coordinate.

The same is true for systems in solution, just substituting the potential energy with the free energy: presently, PCM-like methods have been implemented to compute analytical gradients, while second derivatives can be calculated by numerical methods only. With these procedures minima and saddle points in solution can be effectively located.

Recalling eq. 8.5, the derivative of the free energy with respect to the nuclear coordinate  $\alpha$  can be written

$$(G)^\alpha = (G^{rel})^\alpha + (G^{ster})^\alpha + (G^{dis})^\alpha + (G^{rep})^\alpha \quad (8.44)$$

To compute the various contributions in eq. 8.44 accurately, it is useful to know how the cavity elements change in size and shape when a solute atom moves. In other words one should compute the derivatives of tesseræ areas and positions with respect to nuclear coordinates: at our knowledge, PCM-like methods are the only ones which allow such a calculation [5].

Once these “geometrical” derivatives are known, the gradients of non electrostatic terms are easily computed [5, 34] (see eqs. 8.9 and 8.11):

$$(G^{cav})^\alpha = \sum_i^{spheres} \frac{(A_i)^\alpha}{4\pi R_i^2} G^{cav}(R_i) \quad (8.45)$$

$$\begin{aligned} (G^{dis(rep)})^\alpha &= \rho_{solv} \sum_{s \in S} \sum_{m \in M} \sum_i^{tesseræ} \left[ (a_i)^\alpha V_{ism}^{dis(rep)} \vec{r}_{mi} \cdot \hat{n}_i \right. \\ &\quad \left. + a_i (V_{ism}^{dis(rep)})^\alpha \vec{r}_{mi} \cdot \hat{n}_i + a_i V_{ism}^{dis(rep)} (\vec{r}_{mi} \cdot \hat{n}_i)^\alpha \right] \quad (8.46) \end{aligned}$$

where the variables have been illustrated above. Thus the gradient of the cavitation energy depends on geometrical derivatives only, and also  $(G^{dis(rep)})^\alpha$  is heavily dependent on these quantities, though  $(V_i^{dis(rep)})^\alpha$  and  $(\vec{r}_{mi} \cdot \hat{n}_i)^\alpha$  also depend on the derivatives of nuclear positions. Anyway, all the non electrostatic derivatives need a careful evaluation of geometrical derivatives to be correctly computed.

$(G^{el})^\alpha$  is very often the most important contribution to the free energy gradient in solution: recalling eqs. 8.4 and 8.33 one has

$$\begin{aligned} (G^{el})^\alpha &= \left( \langle \Psi | \hat{H}^0 | \Psi \rangle \right)^\alpha + (V_{NN})^\alpha - tr \mathbf{S}^\alpha (\mathbf{P} \mathbf{F}' \mathbf{P}) \\ &+ \frac{1}{2} tr \mathbf{P} (\mathbf{j}^\alpha + \mathbf{y}^\alpha) + \frac{1}{2} tr \mathbf{P} \mathbf{X}^\alpha (\mathbf{P}) + \frac{1}{2} (U_{NN})^\alpha \end{aligned} \quad (8.47)$$

where  $\left( \langle \Psi | \hat{H}^0 | \Psi \rangle \right)^\alpha + (V_{NN})^\alpha - tr \mathbf{S}^\alpha (\mathbf{P} \mathbf{F}' \mathbf{P})$  is the usual energy gradient calculated as in vacuo (but using the wave function and the Fock matrix perturbed by the solvent). The solvent-dependent terms can be rewritten

$$\begin{aligned} &\frac{1}{2} tr \mathbf{P} (\mathbf{j}^\alpha + \mathbf{y}^\alpha) + \frac{1}{2} tr \mathbf{P} \mathbf{X}^\alpha (\mathbf{P}) + \frac{1}{2} (U_{NN})^\alpha = \\ &= \sum_i^{tesserae} (q_i)^\alpha V_i + \sum_i^{tesserae} q_i (V_i)^\alpha = \quad (8.48) \\ &= \sum_i^{tesserae} (q_i^e)^\alpha V_i^N + \sum_i^{tesserae} q_i^N (V_i^e)^\alpha + \sum_i^{tesserae} (q_i^N)^\alpha V_i^e + \sum_i^{tesserae} q_i^e (V_i^N)^\alpha \\ &+ \sum_i^{tesserae} (q_i^e)^\alpha V_i^e + \sum_i^{tesserae} q_i^e (V_i^e)^\alpha + \sum_i^{tesserae} (q_i^N)^\alpha V_i^N + \sum_i^{tesserae} q_i^N (V_i^N)^\alpha \end{aligned} \quad (8.49)$$

In the formulation 8.49 we have exploited the partition of solvent charges and of solute potential into nuclear and electronic parts. This is the form used in actual calculations: the required derivatives are combinations of the geometrical derivatives mentioned above and of potential and electric field derivatives calculated as in vacuo. The  $(G^{el})^\alpha$  gradients can also be calculated neglecting the geometrical derivatives, i. e. at “fixed cavity”: this has proven to be an effective approximation, sometimes useful to avoid some numerical problems, however in most cases the complete, “mobile cavity” derivatives are preferable.

It is worth noting that the boundary conditions of the COSMO model allows one to simplify eq. 8.49, avoiding the explicit calculation of the solvation charges derivatives, as illustrated in refs. [17], [18] and [31]. This feature makes COSMO a particularly effective model in geometry optimizations. On the other hand, an alternative and very rapid procedure to compute free energy derivatives in the IEF-PCM framework is presently under elaboration.

Once free energy gradients have been determined, they can be used in usual packages for geometry optimizations: the PCM-like procedures implemented in GAUSSIAN98 and GAUSSIAN03 and GAMESS allow geometry optimizations, as well as second derivatives and force constant calculations nearly as effective as in vacuo. Only be obtained numerically, with a considerable computational burden: work is in progress to implement the analytical calculation of the Hessian matrix in PCM-like methods.

Lastly, we note that, even though our analysis has been exclusively devoted to ASC methods [35, 36], this does not mean that the other continuum solvation methods cannot

give the same kind of calculations; indeed analytical expressions for these derivatives are available, for example, for the MPE-SCRF [27] and for the AMSOL methods. The latter, though limited to semi-empirical calculations, allows very effective geometry optimizations in solution, while the former is extended to *ab initio* calculations but it often suffers from convergence problems.

### **b) Numerical applications**

In principle, solvation methods based on effective Hamiltonian approaches allow all the analyses and the manipulations of the solute wave functions that are possible in gas phase. Presently, the most advanced computational procedures, based on the methods illustrated in the preceding chapter, can provide for isotropic solutions:

- the solute molecular free energy, partitioned into electrostatic and non electrostatic terms: the *ab initio* electrostatic contributions can be calculated for closed- and open-shell systems at Hartree-Fock (HF) and Density Functional Theory (DFT) levels, and at some post-HF levels (i.e. MP2, MP4, Coupled Cluster, CI extended to single and double excitations, CAS-SCF);
- the energy gradients with respect to nuclear coordinates, to be used for geometry optimizations in condensed phase, at HF and DFT levels;
- the (numerical) evaluation of harmonic force constants at HF and DFT levels;
- the calculation of first order electronic properties (dipole moments, Mulliken or Bader atomic charges, spin densities and so on) at the same level as the energies (provided the electronic density is available);
- the calculation of static and dynamical electronic polarizability and hyperpolarizabilities, up to the third order, at HF level for closed-shell systems.

The above quantities can be used to study the structure and the reactivity of chemical systems in solution, following the same procedures adopted for isolated molecules. Clearly, the accuracy of the results depend on a number of factors (e.g. the shape of the cavity, the renormalization of the polarization charges, and so on): in the ASC framework (PCM, COSMO and IEF approaches) a good experience exists about these and other computational aspects and, as we shall see, the procedure is particularly flexible and stable.

In the following some examples are reported, using the most advanced implementations of PCM, COSMO and IEF methods (and also mentioning some results obtained by SCRF and AMSOL procedures). The calculations have been performed by the routines implemented by our group in Gaussian94 [37, 38]: we recall that some of the above mentioned quantities can be calculated also by the PCM and IEF routines implemented in Gamess [39] (in particular, electrical polarizability and hyperpolarizabilities in solution

can be computed only by Gamess). DFT calculations have been performed using the mixed functional proposed by Becke, Lee, Yang and Parr (and referred to as B3LYP) [40, 41]; the basis sets are the popular 6-31G [42], sometimes supplemented by diffuse and polarization functions.

PCM-type methods, as well as the most recent versions of SCRF, are based on cavities mimicking the actual molecular shape: usually the well known GEPOL procedure, described in the preceding chapter, is used to build the cavities. As reported in section 8.2 of previous Chapter an alternative cavity definition (called DEFPOL) has been recently proposed.

One of the most important energetic quantity one can get from solvation methods is the solvation free energy

$$\Delta G_{sol} = G - \langle \Psi^0 | \hat{H}^0 | \Psi^0 \rangle \quad (8.50)$$

namely the difference between the solute free energy and the energy of the isolated molecule calculated at the same level (both values are here computed without taking into account entropic terms). In table 1 we report some examples of  $\Delta G_{sol}$  calculated with different solvation models: from the comparison with the corresponding experimental data [43, 44], one can estimate the reliability of this kind of calculations.

**Table 1.**  $\Delta G_{sol}$  (kcal/mol) calculated with some effective Hamiltonian solvation models, compared to experimental results.

Model of solvation	PCM	SCRF	AMSOL	Exp.
Level of calculation	HF	HF	semiempirical	
Basis set	6-31G*	6-31G*		
Kind of cavity	GEPOL with Pauling's radii	ellipsoidal (a)	interlocking spheres	
CH <sub>4</sub>	1.52	0.84		2.0
CH <sub>3</sub> OH	-4.73	-4.88		-5.1
CH <sub>3</sub> CH <sub>2</sub> OH	-3.30	-4.07		-5.0
CH <sub>3</sub> COOH	-7.50	-6.72	-6.6	-6.7
CH <sub>3</sub> CONH <sub>2</sub>	-10.21	-9.71		-9.7
CH <sub>3</sub> COCH <sub>3</sub>	-3.50	-4.3		-3.8
CH <sub>3</sub> CHO	-4.59	-3.44		-3.5
CH <sub>3</sub> NH <sub>2</sub>	-3.13	-4.87		-4.6
CH <sub>3</sub> CN	-5.66	-3.97		-3.9
H <sub>3</sub> O <sup>+</sup>	-94.3		-109.4	-104
CH <sub>3</sub> OH <sub>2</sub> <sup>+</sup>	-77.0		-89.7	-85
(CH <sub>3</sub> ) <sub>2</sub> COH <sup>+</sup>	-60.0		-65.7	-64
CH <sub>3</sub> NH <sub>3</sub> <sup>+</sup>	-74.0		-78.4	-70
OH <sup>-</sup>	-97.8		-111.8	-106
Ph-O <sup>-</sup>	-63.5		-68.3	-72

(a) non electrostatic contributions computed with GEPOL-like cavities and parameterized to fit experimental data in ref.[45].

**Table 2.**  $\Delta G_{sol}$  (kcal/mol) calculated by PCM at HF level (basis set: 6-31G\* for neutrals, 6-31+G\* for ions) with UATM cavity model.

	Calculated <sup>(a)</sup>	Exp.		Calculated <sup>(a)</sup>	Exp.
	$\Delta G_{sol}$	$\Delta G_{sol}$		$\Delta G_{sol}$	$\Delta G_{sol}$
CH <sub>4</sub>	1.86	2.0	H <sub>3</sub> O <sup>+</sup>	-105.3	-104
H <sub>2</sub> O	-6.28	-6.3	CH <sub>3</sub> OH <sub>2</sub> <sup>+</sup>	-87.4	-85
CH <sub>3</sub> OH	-4.99	-5.1	(CH <sub>3</sub> ) <sub>2</sub> COH <sup>+</sup>	-65.2	-64
(CH <sub>2</sub> OH) <sub>2</sub>	-9.80	-9.6	NH <sub>4</sub> <sup>+</sup>	-79.8	-79
NH <sub>3</sub>	-4.34	-4.3	CH <sub>3</sub> NH <sub>3</sub> <sup>+</sup>	-70.3	-70
CH <sub>3</sub> NH <sub>2</sub>	-4.57	-4.6	Ph-NH <sub>3</sub> <sup>+</sup>	-64.1	-66
Ph-NH <sub>2</sub>	-4.69	-4.9	CH <sub>3</sub> SH <sub>2</sub> <sup>+</sup>	-75.6	-74
CH <sub>3</sub> COOH	-7.10	-6.7	OH <sup>-</sup>	-108.1	-106
CH <sub>3</sub> CONH <sub>2</sub>	-9.40	-9.7	CH <sub>3</sub> O <sup>-</sup>	-94.3	-95
Ph-NO <sub>2</sub>	-3.97	-4.1	CH <sub>3</sub> COO <sup>-</sup>	-78.1	-77
pyridine	-4.79	-4.7	Ph-O <sup>-</sup>	-71.3	-72
NH <sub>2</sub> -NH <sub>2</sub>	-9.65	-9.3	PH <sub>2</sub> <sup>-</sup>	-65.9	-65

(a) renormalization method ICOMP=4 (see below).

The PCM average error can be reduced by one order of magnitude by employing more refined definition of cavities, as shown in table 2.

The UATM (United Atom Topological Model) cavities are based on the GEPOLE procedure, the main difference being that hydrogen atoms are always included in the same sphere as the heavy atom they are bonded to. The radius of each atomic or group sphere is defined in terms of the molecular topology: with a limited number of parameters, UATM cavities allow one to calculate  $\Delta G_{sol}$  with chemical accuracy for a wide range of compounds (on a set of 43 neutral solutes, the average error is lower than 0.2 kcal/mol, whereas on 27 ions an average error of about 1 kcal/mol is found) [46].

The dependence of the results on the charge renormalization in the framework of PCM, COSMO and IEF methods can be judged from the results shown in table 3. Notice that a deep analysis of the numerical errors due to the discretization and to the escaped electronic tails has been performed for ASC methods (PCM, COSMO, IEF) only. In table 3 the three different renormalization procedures introduced in the previous chapter (see section 8.3.3) have been indicated as follows. ICOMP=0 means that no normalization has been applied (this option is effective for COSMO and IEF models only, whereas with PCM it gives unreliable results); ICOMP=2 means that each charge has been multiplied by the same factor  $f^x$  in order to bring the total polarization charge to the value predicted by Gauss' law. ICOMP=3 indicates that the charges have been renormalized tessera by tessera with a correction proportional to the solute electronic density; lastly, ICOMP=4 (active for PCM calculations only) corresponds to the most complex method, taking into account the escaped charge by an effective polarization charge spread on the cavity surface (see refs. [29] and [30] for a more comprehensive review of normalization methods adopted in PCM).



**Table 3.**  $\Delta G_{sol}$  (kcal/mol) computed by PCM, COSMO and IEF solvation models with different renormalizations (ICOMP) of the solvation charges

	PCM			COSMO			IEF			Exp.
	2	3	4	0	2	3	0	2	3	
H <sub>2</sub> O	-6.53	-6.38	-6.29	-6.26	-6.26	-6.36	-6.20	-6.20	-6.19	-6.3
CH <sub>3</sub> OH	-5.32	-5.10	-5.99	-5.04	-5.05	-5.12	-4.99	-5.00	-4.98	-5.1
CH <sub>3</sub> NH <sub>2</sub>	-5.04	-4.81	-4.57	-4.69	-4.68	-4.75	-4.65	-4.64	-4.62	-4.6
pyridine	-5.23	-5.35	-4.78	-4.66	-4.68	-4.74	-4.62	-4.63	-4.74	-4.7
CH <sub>3</sub> OH <sub>2</sub> <sup>+</sup>	-89.3	-87.9	-87.4	-88.8	-88.6	-89.3	-88.6	-88.4	-88.3	-85
OH <sup>-</sup>	-111.6	-109.8	-108.1	-104.2	-111.0	-110.7	-104.5	-110.8	-110.5	-106
Ph-S <sup>-</sup>	-58.9	-63.9	-65.4	-62.7	-64.9	-65.1	-92.9	-64.8	-65.2	-67

It is noteworthy that COSMO and IEF models need a simpler normalization procedure (say ICOMP=2) to give results of accuracy comparable to that of PCM with the most advanced procedure (namely ICOMP=4).

Though multipole expansion methods like SCRf are based on an approximate description of the electrostatic potential, their major advantage is that the solvent effects can be calculated very quickly, mainly with simple spherical or ellipsoidal cavities. On the other hand, in many cases it is necessary to reach high multipole orders to approach convergence.

Such convergence problems are avoided by ASC methods: furthermore, the most recent implementations of PCM, COSMO and IEF in Gaussian94 have reached a noticeable degree of computational efficiency.

From the comparison of the CPU time needed for some HF calculations *in vacuo* and in solution (by PCM and COSMO), it is clear that a SCF calculation taking into account solvent effects requires less than twice the time needed for an isolated molecule. Notice that in most tests, the SCF in solution is composed by two successive calculations, the first referred to the isolated solute, the second taking into account the solvent effects: this way, one obtains a number of information with a single run: however, it is possible to perform the calculation in the solvent directly, saving more time.

When free energy gradients are calculated in solution, one generally finds that the electrostatic contribution ( $G_{el}^{\alpha}$ , see eqs. 8.44 and 8.47) is the dominant one. For example, in table 4 we report the various contributions to the largest gradient components for some molecules in aqueous solution (notice that the geometry of all these systems had previously been optimized in gas phase, so that all the gradient components would be lower than 0.00045 a. u. in the absence of the solvent). Of course, in the case of large, flexible and non polar compounds the contributions on non electrostatic (mainly cavitation) energy gradients can become crucial.

**Table 4.** Electrostatic and non electrostatic contributions (a. u.) to free energy gradients. All molecules at the geometries optimized *in vacuo* at the same level.

	Level of calculation	Radii	Largest <sup>(a)</sup>	Largest <sup>(a)</sup>	Largest <sup>(a)</sup>	Largest <sup>(a)</sup>
			$\Delta G_{el}^\alpha$ (a.u.)	$\Delta G_{cav}^\alpha$	$\Delta G_{dis}^\alpha$	$\Delta G_{rep}^\alpha$
H <sub>2</sub> O	HF/6-31G*	Pauling	0.01518	0.00096	0.00030	0.00008
H <sub>3</sub> O <sup>+</sup>	HF/6-31G*	Pauling	0.09976	0.00091	0.00039	0.00007
OH <sup>-</sup>	HF/6-31+G*	Pauling	0.04471	0.00095	0.00027	0.00008
CH <sub>3</sub> OH	HF/6-31G*	UATM	0.02175	0.00107	0.00087	0.00017
CH <sub>3</sub> O <sup>-</sup>	HF/6-31+G*	UATM	0.09096	0.00094	0.00113	0.00175
CH <sub>3</sub> CONH <sub>2</sub>	HF/6-31G*	UATM	0.02284	0.00111	0.00104	0.00158
CH <sub>3</sub> COOH	B3LYP/6-31G**	UATM	0.03306	0.00113	0.00087	0.00019

(a) In the optimization algorithm implemented in Gaussian94 the standard convergence threshold on gradients is 0.00045 a. u.

The availability of reliable energy gradients is going to make the geometry optimization of systems in solution as easy and fast as for isolated molecules.

Another interesting result is reported in table 5, where we list the geometrical parameters of two conformers of the glycine radical zwitterion (<sup>+</sup>H<sub>3</sub>N- $\dot{C}$ H-COO<sup>-</sup>):

optimized *in vacuo* and in water by PCM: it is noteworthy that the conformer 2, with an aminic hydrogen in *cis* position with respect to an oxygen, is not a stationary point *in vacuo*, whereas a minimum in the free energy hypersurface appears in solution.

**Table 5.** Geometrical parameters (bond lengths in Å and valence angles in degrees) for two conformers of the zwitterionic form of glycine radical. Calculations at B3LYP/6-31+G\*\* level.

	Conformer 1 <i>in vacuo</i>	Conformer 1 in water	Conformer 2 in water
NC	1.486	1.485	1.451
CC	1.500	1.481	1.482
CO <sub>c</sub>	1.271	1.268	1.272
CO <sub>t</sub>	1.245	1.264	1.262
CH	1.084	1.082	1.081
NH <sub>1</sub>	1.023	1.026	1.037
NH <sub>2</sub>	1.031	1.029	1.029
NCC	110.87	118.13	115.91
CCO <sub>c</sub>	111.44	116.11	115.11
CCO <sub>t</sub>	115.78	115.80	116.76
CCH	132.11	126.26	126.58
CNH <sub>1</sub>	115.19	111.62	106.03
CNH <sub>2</sub>	107.31	111.81	112.83
H <sub>2</sub> NCC	54.42	59.00	119.68

Also other methods referenced in the preceding chapter are able to perform geometry optimizations in solution: for example, in the semi classical framework, the AMSOL procedure has been parameterized for a great variety of compounds and it can calculate energy derivatives very quickly. Besides PCM-like methods, another widely used *ab initio* solvent model providing good geometry optimizations is SCRF; moreover, other implementations of COSMO, different from ours, have been proposed for the *ab initio* optimizations in solution. Unlike the other *ab initio* models, the PCM and the COSMO procedures implemented by our groups can calculate energy gradients taking into account the non electrostatic terms, and also the effects due to the deformation of the cavity. As for single point calculations, the most effective implementations require a computational effort only slightly greater in solution than *in vacuo*

According to the Transition State Theory, a chemical reaction can be represented on the suitable energy hypersurface as the minimum energy path connecting two minima (reactants and products) through a first order saddle point (transition state). Then, a further step in the analysis of energy hypersurfaces is the location of saddle points, that can be accomplished in solution with the help of the methods elaborated for isolated systems, though the convergence of such search algorithms sometimes raises some problems when the solvent effects are taken into account.

An example of transition state optimization is shown in table 6, where we report some data for the interconversion between formamide and formamidinic acid assisted by one water molecule:

**Table 6.** Geometries, bond lengths in Å and valence angles in degrees, and energy differences (kcal/mol) of minima and transition state for the interconversion formamide  $\rightarrow$  formamidinic acid assisted by one water molecule. Calculations *in vacuo* and in solution at B3LYP/6-31G\*\* level.

	Formamide		Transition state		Formamidinic acid	
	<i>in vacuo</i>	in water	<i>in vacuo</i>	in water	<i>in vacuo</i>	in water
Energy	0.0	0.0	19.4	20.8	10.1	12.4
NH <sub>1</sub>	1.022	1.017	1.318	1.335	2.434	1.930
NH <sub>2</sub>	1.007	1.010	1.012	1.011	1.017	1.017
NC	1.348	1.333	1.308	1.303	1.280	1.276
CO	1.230	1.243	1.284	1.295	1.328	1.338
CH	1.105	1.100	1.097	1.094	1.095	1.092
OH <sub>w</sub>	1.899	1.852	1.220	1.273	0.999	0.996
OO <sub>w</sub>	2.800	2.794	2.397	2.401	2.611	2.708
O <sub>w</sub> H <sub>w</sub>	0.966	0.967	0.967	0.970	0.966	0.969
O <sub>w</sub> H <sub>1</sub>	1.975	2.075	1.192	1.175	0.988	0.986
H <sub>1</sub> NC	117.6	118.5	105.3	105.9	55.6	104.4
H <sub>2</sub> NC	121.0	120.8	116.6	116.1	111.4	109.2
OCN	124.9	124.3	122.0	121.6	123.5	123.4
HCN	113.8	114.4	121.3	122.3	125.2	125.1
H <sub>w</sub> OC	105.6	105.2	103.2	102.3	111.3	107.9
H <sub>w</sub> O <sub>w</sub> H <sub>w</sub>	104.3	107.2	109.5	107.1	104.9	103.8
H <sub>w</sub> O <sub>w</sub> H <sub>w</sub> O	112.0	101.0	108.2	106.4	109.8	104.3

The knowledge of minima and saddle points geometry and energy allows one to determine activation energies and to get many insights about reaction mechanisms: the next step would be the calculation of reaction coordinates (for example the Intrinsic Reaction Coordinate, IRC, proposed by Fukui). However, the determination of IRC's for complex systems requires a good estimate of the Hessian matrix, at least in the saddle point: presently the PCM-type methods we have illustrated above do not provide analytical energy second derivatives in solution, and numerical estimates are not accurate enough for this scope. Other approaches, mainly the SCRF, are able to calculate Hessian matrices in solution: however, many convergence problems are encountered, so that one can say that more work is still needed to describe chemical reactions in solution at the same level as *in vacuo*.

On the other hand, the numerical evaluation of free energy second derivatives is accurate enough to calculate vibrational frequencies in solution (in this case attention must be paid to the shape of the cavity: for example, united atom models seem less accurate than descriptions with a sphere around each atom). In table 7 we report harmonic wave numbers and IR intensities calculated for formamide in water. Inclusion of solvent effects substantially improves the agreement with the experiments: of course the numerical derivation is cumbersome, and the calculation in solution is much longer than *in vacuo*.

**Table 7.** Harmonic wave numbers ( $\text{cm}^{-1}$ ) and IR intensities ( $\text{km/mol}$ ) for formamide calculated *in vacuo* and in water (by PCM) at B3LYP/6-31G\* level.

Vibrational mode	Wave numbers			IR intensities		
	<i>in vacuo</i>	in water	exp.	<i>in vacuo</i>	in water	exp.
NH <sub>2</sub> inversion	162.1	365.6	303	249	345	very strong
NCO bend	562.8	581.2	564	12	26	medium
NH <sub>2</sub> twist	649.6	681.8	682	21	16	medium
CH out of plane	1047.2	1070.5	1047	0	1	weak
NH <sub>2</sub> wagging	1054.4	1070.7	1185	3	5	weak
CN stretch	1275.5	1301.6	1261	95	111	strong
CH in plane	1432.5	1405.1	1400	54	30	medium
NH <sub>2</sub> bend	1620.8	1588.1	1579	68	119	strong
CO stretch	1836.0	1746.2	1740	352	651	ultra strong
CH stretch	2949.2	3012.0	2884	110	89	medium
NH <sub>2</sub> symm. str.	3596.6	3574.8	3427	314	123	medium
NH <sub>2</sub> asymm. str.	3741.2	3713.8	3548	37	130	medium

**Table 8.** Mulliken group charges (a.u.) calculated *in vacuo* and in two solvents at HF/6-31G\* level.

	Charge on fragment	<i>In vacuo</i>	In chloroform	In water
CH <sub>3</sub> OH	-OH	-0.288	-0.301	-0.306
CH <sub>3</sub> CH <sub>2</sub> OH	-OH	-0.296	-0.308	-0.313
CH <sub>3</sub> NH <sub>2</sub>	-NH <sub>2</sub>	-0.168	-0.175	-0.178
Ph-OH	-OH	-0.300	-0.297	-0.294
Ph-NH <sub>2</sub>	-NH <sub>2</sub>	-0.206	-0.173	-0.159
Ph-SH	-SH	+0.112	+0.108	+0.108

As said at the beginning of this chapter, effective Hamiltonian solvation methods produce solute wave functions (expressed as density matrices on atomic bases) that can be analyzed exactly as the corresponding wave functions for isolated molecules. For example Mulliken population analysis, widely used to get a rough but useful description of electronic distributions, can be applied to molecular solutes, as shown e.g. in table 8. Of course, more refined analyses can be performed, too: for example, Bader's procedure can be used to partition the electronic properties (like the charge) between atoms, at a computational cost exactly equal *in vacuo* and in solution.

Dipole moments can be computed in solution as well as *in vacuo*: in table 9 we report the dependence of this quantity on the calculation level for some test molecules.

**Table 9.** Dipole moment (Debye) calculated *in vacuo* and in water (PCM) at different calculation levels (basis set: 6-31G\*).

	H <sub>2</sub> O		CH <sub>3</sub> OH	
	<i>in vacuo</i>	in water	<i>in vacuo</i>	in water
HF	2.186	2.479	1.857	2.210
MP2 <sup>(a)</sup>	2.151	2.451	1.762	2.132
MP4 <sup>(a)</sup>	2.135	2.435	1.764	2.132
QCISD <sup>(a)</sup>	2.143	2.441	1.791	2.156
CCSD <sup>(a)</sup>	2.140	2.438	1.777	2.143
B3LYP	2.050	2.345	1.645	1.980

(a) Using solvation charges determined at HF level.

As said in Chapter 1, one usually calls “direct” solvent effects those related to the polarization of the solute wave function due to the solvent reaction field, keeping the geometry optimized in *vacuo*, while the modifications due to the geometry relaxation induced by the solvent are referred to as “indirect” effects. It is noteworthy that the procedures presented above are able to separate direct and indirect solvent effects on molecular energies and structures.

A final note: this presentation was written some time ago. In the meanwhile some important aspects of the PCM model as well as many implementative issues have been

changed and updated: the reader can find an illustration of the most recent developments in refs. [47], [48] and references therein.

# Bibliography

- [1] J.B. Foresman, T.A. Keith, K.B. Wiberg, J. Snoonian, M.J. Frisch, *J. Phys. Chem.* 100, 16098 (1996).
- [2] J.L. Pascual-Ahuir, E. Silla, J. Tomasi and R. Bonaccorsi, *J. Comp. Chem.* 8, 778 (1987).
- [3] R. Cammi, M. Cossi, B. Mennucci, C.S. Pomelli and J. Tomasi, *Int. J. Quantum Chem.* 60, 1165 (1996).
- [4] B. Lee and F.M. Richards, *J. Mol. Biol.* 55, 379 (1971).
- [5] M. Cossi, M. Mennucci and R. Cammi, *J. Comp. Chem.* 17, 57 (1996).
- [6] H.H. Uhlig, *J. Phys. Chem.* 41, 1215 (1937).
- [7] O. Sinanoglu, *J. Chem. Phys.* 75, 463 (1981).
- [8] V. Gogonea and E. Osawa, *J. Mol. Struct. (Theochem)* 311, 305 (1994).
- [9] H. Reiss, H.L. Frisch, J.L. Lebowitz, *J. Chem. Phys.* 31, 369 (1959).
- [10] R. A. Pierotti, *Chem. Rev.* 76, 717 (1976).
- [11] F.M. Floris and J. Tomasi, *J. Comput. Chem.* 10, 616 (1989); F.M. Floris, J. Tomasi and J.L. Pascual-Ahuir, *J. Comput. Chem.* 12, 784 (1991).
- [12] J. Langlet, P. Claverie, J. Caillet and A. Pullman, *J. Phys. Chem.* 92, 1617 (1988).
- [13] C. Amovilli and B. Mennucci, *J. Phys. Chem B* 101, 1051 (1997)
- [14] S. Miertuš, E. Scrocco and J. Tomasi, *Chem. Phys.* 55, 117 (1981).
- [15] J. Tomasi and M. Persico, *Chem. Rev.* 94, 2027 (1994).
- [16] R. Cammi and J. Tomasi, *J. Comp. Chem.* 16, 1449 (1995).
- [17] A. Klamt and G. Schüürmann, *J. Chem. Soc. Perkin Trans. 2*, 799 (1993)
- [18] T. N. Truong and E.V. Stefanovich, *Chem. Phys. Lett.* 240, 253 (1995).
- [19] E. Cancès and B. Mennucci, *J. Math. Chem.*, 23, 309 (1997).

- [20] E. Cancès, B. Mennucci and J. Tomasi, *J. Chem. Phys.* 107, 3032 (1997).
- [21] D. Rinaldi, M.F. Ruiz-López and J.L. Rivail, *J. Chem. Phys.* 78, 834 (1983).
- [22] V. Dillet, D. Rinaldi and J.L. Rivail, *J. Phys. Chem.* 98, 5034 (1994).
- [23] C.J. Cramer, G.C. Lynch, G.D. Hawking, D.G. Truhlar, *QCPE Bull: AMSOL version 4.0* 13, 78 (1993).
- [24] M. Born, *Z. Phys.* 1, 45 (1920).
- [25] C.J. Cramer and D.G. Truhlar, *Reviews on Computational Chemistry*, D.B. Boyd and K.B. Lipkowitz (eds.), VCH, New York (1995); Vol. 6.
- [26] J.G. Kirkwood, *J. Chem. Phys.* 2, 351 (1934).
- [27] V. Dillet, D. Rinaldi, J. Bertrán and J.L. Rivail, *J. Chem. Phys.* 104, 9437 (1996).
- [28] D. Rinaldi, B.J. Costa-Cabral and J.L. Rivail, *Chem. Phys. Lett.* 125, 495 (1986).
- [29] B. Mennucci and J. Tomasi, *J. Chem. Phys.* 106, 5151 (1997).
- [30] M. Cossi, B. Mennucci, J. Tomasi and J. Pitarch, *J. Comp. Chem.* in press
- [31] V. Barone and M. Cossi, *J. Phys. Chem.* in press
- [32] A. Klamt and V. Jonas, *J. Chem. Phys.* 105, 9972 (1996).
- [33] W. Hackbusch, *Integral equations—Theory and numerical treatment*, Birkhäuser Verlag, Basel, Switzerland, 1995.
- [34] M. Cossi, J. Tomasi and R. Cammi, *Int. J. Quantum Chem.: Quantum Chem. Symp.* 29, 695 (1995).
- [35] R. Cammi and J. Tomasi, *J. Chem. Phys.* 101, 3888 (1994).
- [36] T. N. Truong and E.V. Stefanovich, *J. Chem. Phys.* 103, 3709 (1995).
- [37] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill P. M. W.; Johnson B. G.; Robb, M. A.; Cheeseman J. R.; Keith, T.; Petersson G. A.; Montgomery, J.A.; Raghavachari, K.; Al-Laham, M.A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J.B.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94, Revision D.4* Gaussian, Inc., Pittsburgh PA (1995).
- [38] M. Cossi, V. Barone, R. Cammi and J. Tomasi, *Chem. Phys. Lett.*, 225, 327, (1996).
- [39] M.W. Schmidt, K.K. Baldrige, J.A. Boatz, S.T. Elbert, M.S. Gordon, J.H. Jensen, S. Koseki, N. Matsunaga, K.A. Nguyen, S.J. Su, T.L. Windus, M. Dupuis and J.A. Montgomery, *J. Comp. Chem.* 14, 1347 (1993).
- [40] A. D. Becke, *Phys. Rev. B* 38, 3098 (1988).



- [41] C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B* 37, 785 (1988).
- [42] P. C. Hariharan and J. A. Pople, *Theor. Chim. Acta* 28, 213 (1973).
- [43] S. Cabani, G. Mollica and V. Lepori, *J. Sol. Chem.* 10, 563 (1981).
- [44] R. G. Pearson, *J. Am. Chem. Soc.* 108, 6109 (1986).
- [45] I. Tuñón, M. Ruiz-Lopez, D. Rinaldi and J. Bertrán, *J. Comput. Chem.* 17, 148 (1996).
- [46] V. Barone, M. Cossi and J. Tomasi, *J. Chem. Phys.* 107, 3210 (1997).
- [47] J. Tomasi, R. Cammi, B. Mennucci, C. Cappelli and S. Corni, *Phys. Chem. Chem. Phys.* 4, 5697 (2002).
- [48] M. Cossi, G. Scalmani, N. Rega and V. Barone, *J. Chem. Phys.* 117, 43 (2002).