Research paper

Analysis of solute-solvent interactions using the solvation model density combined with the fragment molecular orbital method

Dmitri G. Fedorov

Research Center for Computational Design of Advanced Functional Materials (CD-FMat), National Institute of Advanced Industrial Science and Technology (AIST), Central 2, Umezono 1-1-1, Tsukuba 305-8568, Japan

ARTICLE INFO

ABSTRACT

The energy and its analytic gradient are developed for the solvation model density (SMD) combined with the fragment molecular orbital (FMO) method. The accuracy of the energy is evaluated in comparison to full results without fragmentation for a set of neutral and charged polypeptides. The accuracy of the gradient is computed in comparison to numerical gradient. The components to the solvation energy in SMD are compared to the polarizable continuum model (PCM). FMO with SMD and PCM is applied to analyze protein–ligand binding in solution.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Many chemical processes occur in solution, and it is necessary to take into account solvent effects [1] in molecular simulations. With an explicit treatment of solvent one needs to do long time dynamics simulations with many solvent molecules, which makes this route very costly for practical applications. An alternative, often employed in quantum-chemical (QM) calculations, is to use a continuum solvation model, a compromise between accuracy and computational efficiency.

QM calculations of large molecular systems [2,3] can be done with linear-scaling [4,5], fragment-based [6–16] or other approaches [17–22]. In the fragment molecular orbital (FMO) method [23–27], a molecular system is divided into fragments, and their electronic state is computed in the presence of an embedding electrostatic potential (ESP), dependent on the electron densities of fragments (monomers). After convergence of monomers, fragment pair (dimer) calculations are performed, and finally the total properties are calculated using the many-body expansion [28]. To increase the accuracy, trimer calculations can be performed [29]. FMO has been combined with the polarizable continuum model (PCM) [30–32] and other solvation approaches [33–35].

In this work, FMO is combined with the solution model density (SMD) [36] up to the three-body level. Although FMO/PCM has been successfully applied to many systems, it has been argued [37,38] that PCM can give poor results for some protein cavities, and thus it is worth having an alternative to PCM. In addition, SMD has a systematically defined set of radii for most atoms, whereas some components of PCM lack parameters for some

atoms. Solvent effects are discussed comparing PCM and SMD, and both methods are applied to biochemical systems to discern their similarities and differences.

2. Methodology

SMD has two components, electrostatic and non-polar [36]. The former is evaluated as in PCM, but numerically the results differ because the two models use different atomic radii. The solute is placed into a cavity, made of a union of atomic spheres, and each sphere is divided into pieces called tesserae, with a point charge for each tessera. These solvent charges are polarized (induced) by the solute and in turn they polarize the solute (a self-consistent mutual polarization). The non-polar component in SMD combines cavitation, dispersion, and solvent structure effects (cds). This cds term corresponds to cavitation, dispersion, and repulsion interactions (cdr) in PCM. The cds and cdr terms are parametrized in an entirely different way, but in either case the parametrization employed in this work is independent of the electronic state of the solute.

The FMO energy expression at the two-body level (FMO2) for a system divided into N fragments is

$$E = \sum_I^N \left(E_I'' + \Delta E_I^{\text{solv}} \right) + \sum_{I>J}^N \Delta E_{IJ}^{\text{int}} \qquad (1)$$

where the internal solute E_I'' and solvent-solute interaction energies ΔE_I^{solv} of fragments I are added to the interaction (int) energies $\Delta E_{IJ}^{\text{int}}$ between fragments I and J,

$$\Delta E_{IJ}^{\text{int}} = \left(E_{IJ}'' - E_I'' - E_J'' \right) + Tr\left(\Delta \mathbf{D}^{IJ} \mathbf{V}^{IJ} \right) + \Delta E_{IJ}^{\text{solv}} \qquad (2)$$

E-mail address: d.g.fedorov@aist.go.jp

$\Delta\mathbf{D}^{IJ}$ is the electron density difference (dimer minus two monomers), and $\mathbf{V}^{IJ}$ is the embedding solute ESP for dimer IJ due to fragments other than I and J. ΔE_{IJ}^{solv} is the solvent screening of the I-J interaction. In this work, explicit equations are given for FMO2, whereas the corresponding equations for the three-body FMO (FMO3) are omitted for simplicity.

In PCM, the solvent-related terms are

$$\Delta E_I^{solv} = \Delta E_I^{es} + \Delta E_I^{cdr} \tag{3}$$

$$\Delta E_{IJ}^{solv} = \Delta E_{IJ}^{es} + \Delta E_{IJ}^{CT\cdot es} + \Delta E_{IJ}^{dr} \tag{4}$$

The dimer term ΔE_{IJ}^{solv} has an electrostatic (es) and dispersion-repulsion (dr) components, as well as the coupling of the charge transfer (CT) between fragments I and J to the embedding potential due to solvent (CT · es). In SMD, the corresponding terms are

$$\Delta E_I^{solv} = \Delta E_I^{es} + \Delta E_I^{cds} \tag{5}$$

$$\Delta E_{IJ}^{solv} = \Delta E_{IJ}^{es} + \Delta E_{IJ}^{CT\cdot es} \tag{6}$$

Using pair interaction energy (PIE) decomposition analysis (PIEDA) [39–41], it is possible to decompose ΔE_{IJ}^{int} into the electrostatic, exchange-repulsion, charge transfer with mix terms, DFT correlation, dispersion and solvent screening. A significant difference between PCM and SMD is in the monomer terms, ΔE_I^{cdr} (PCM) vs ΔE_I^{cds} (SMD). PIEs also differ because of the absence of ΔE_{IJ}^{dr} in SMD, which typically contribute relatively little (at most 1–2 kcal/mol), dominated by the solute-solvent dispersion. The values of $\Delta E_{IJ}^{es} + \Delta E_{IJ}^{CT\cdot es}$ are not the same in SMD and PCM because different radii are used.

There are several levels in FMO/PCM for the embedding solvent potential $\widetilde{\mathbf{V}}$ computed from the fragment contributions $\widetilde{\mathbf{V}}^I$ and $\widetilde{\mathbf{V}}^{IJ}$.

$$\widetilde{\mathbf{V}} = \sum_I^N \widetilde{\mathbf{V}}^I + c_V \sum_{I>J}^N \left(\widetilde{\mathbf{V}}^{IJ} - \widetilde{\mathbf{V}}^I - \widetilde{\mathbf{V}}^J \right) \tag{7}$$

In the more computationally efficient methods PCM[1] and PCM⟨1⟩, $c_V = 0$, whereas in the more accurate PCM[1(2)] and PCM[2], $c_V = 1$ (see elsewhere [42] for a description of these levels). Because SMD shares the electrostatic solvent treatment and potential with PCM, the same gradation of levels can be done with SMD. In this work, both PCM and SMD are used at the level of ⟨1⟩ [32].

The FMO/SMD gradient is evaluated by taking the derivative of the total energy in Eq. (1) with respect to a nuclear coordinate a ($\partial E / \partial a$). The self-consistent Z-vector (SCZV) method [43,32] is used in SMD in the same way as in PCM. The gradient of the cds term, which does not depend on the electronic state, is evaluated separately also analytically [36].

In the subsystem analysis [44], the binding energy ΔE is decomposed into the contributions of fragments. The subsystem analysis previously developed for PCM is extended to SMD.

$$\Delta E = \sum_I^N \Delta E_I^{bind} \tag{8}$$

The fragment binding energies ΔE_I^{bind} include the desolvation and polarization effects of fragments, and residue-ligand interactions. The deformation contributions are also implicitly included by taking structures separately obtained for the bound and isolated states. The sum in Eq. (8) runs over all fragments in the complex, i.e., all residues and the ligand.

For the purpose of analysis, it is useful to decompose the total energy in Eq. (1) into the solute energy E' and the solvent-solute interaction energy ΔE^{solv}.

$$E = E' + \Delta E^{solv}$$
$$E' = \sum_I^N E_I'' + \sum_{I>J}^N \left[\left(E_{IJ}'' - E_I'' - E_J'' \right) + Tr\left(\Delta\mathbf{D}^{IJ}\mathbf{V}^{IJ} \right) \right] \tag{9}$$
$$\Delta E^{solv} = \sum_I^N \Delta E_I^{solv} + \sum_{I>J}^N \Delta E_{IJ}^{solv} = \Delta E^{es} + \Delta E^{np}$$

where the non-polar (np) contribution ΔE^{np} for PCM is ΔE^{cdr}, and for SMD, ΔE^{cds} (none of the two includes the additional contribution of 1.89 kcal/mol to accommodate for the change from 1 atm gas to 1 M solution [45]).

3. Computational details

For all calculations, FMO implemented in GAMESS [46] was used in parallel [47]. The computation of SMD terms was parallelized. Hybrid orbital projection operators were used to describe fragment boundaries. The solvent (water) was treated using C-PCM or SMD using the FIXPVA [48] tesselation with 60 tesserae per atom. In PCM, SUAHF radii were used as recommended for FIXPVA; for SMD, its own unique set of radii [36] was used. The polypeptides were divided as 1 residue per fragment at C atoms; ligand was treated as a separate fragment. In DFT, the SG-1 grid and D3(BJ) dispersion [49] were used. Geometry optimizations were done with the threshold OPTTOL = 10^{-4} hartree/bohr.

The numerical tests are as follows. First, the accuracy of FMO/SMD is evaluated with respect to the full SMD (full in this work means unfragmented QM without using FMO) for a set of neutral and charged capped α-helices $(ALA)_nX$, where X is ALA (neutral), ARG (cation) and GLU (anion) and $n = 9, 19$ and 39. For comparison, similar tests were also done for PCM. The structures for these systems, denoted as α-$(ALA)_nX$ were taken from the previous study [31]. The results reported in this Letter for density functional theory can be compared to the accuracy of restricted Hartree-Fock
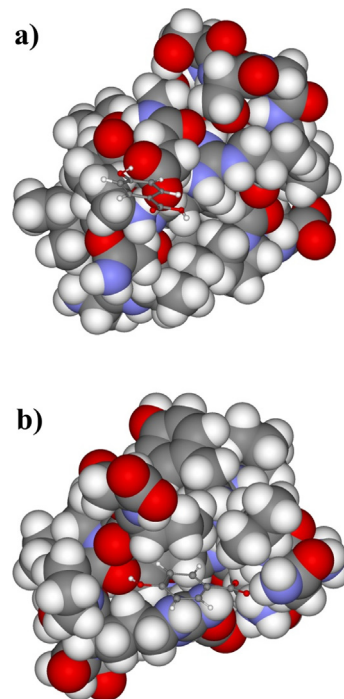

Fig. 1. Structures of the complexes of the Trp-cage protein and (a) neutral and (b) anionic ligands.