Electronic Supplementary Information for

Generalized energy-based fragmentation approach for calculations

of solvation energies of large systems

Kang Liao, Shirong Wang, Wei Li* and Shuhua Li*

School of Chemistry and Chemical Engineering, Key Laboratory of Mesoscopic Chemistry of Ministry of Education, Institute of Theoretical and Computational Chemistry, Nanjing University, Nanjing, 210023, P. R. China.

E-mail: wli@nju.edu.cn, shuhua@nju.edu.cn

1. The Derivation of GEBF-SMD Electronic-Polarization Energy

We will derive the GEBF-SMD electronic-polarization (EP) energy of a target system in solution (with solute cavity), as displayed in Fig. S1(I). In all the subsequent equations, the cavitation-dispersion-solvent-reorganization (CDS) free energy term will not be included.
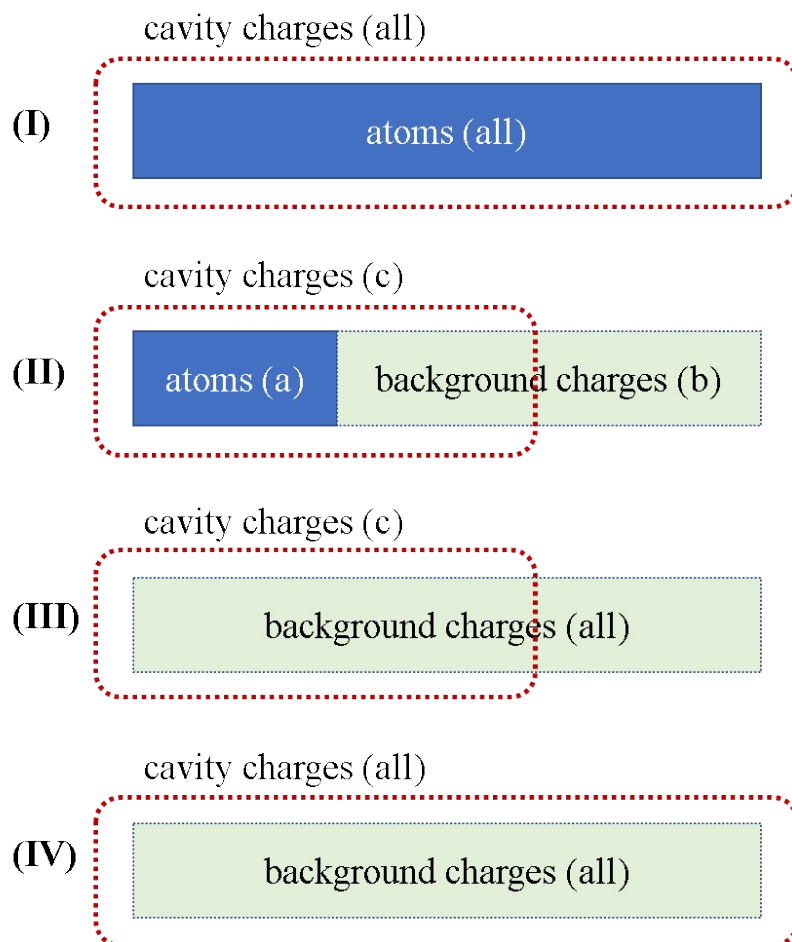
Fig S1. An illustration of systems to be calculated in the GEBF-SMD approach: (I) The entire system with full solute cavity; (II) An embedded subsystem ($^S m$) with truncated cavity; (III) A pure charge system ($^S (m)$) with truncated cavity, which is obtained by replacing the atoms with charges in $^S m$; (IV) A pure charge system with full background point charges and full cavity. The atoms and background point charges are denoted as blue and light green rectangle, and the surface of solute cavity is denoted as dotted red box.

Firstly, the GEBF-SMD EP energy of the entire system in solution (I), i.e. Eq. (8) in the main text, can be approximately expressed as,

$$E_{EP}^{GEBF} = \sum_m C_m \left(E_m^{sol} - E_{(m)}^{sol} \right) + E_{(all)}^{sol}$$

(S1)

Here E_m^{sol} is the total energy of the mth embedded subsystem (S_m) with truncated cavity [see Fig. S1 (II)], $E_{(m)}^{sol}$ is the total energy of one pure charge system $S_{(m)}$ [see Fig. S1(III)] with full background point charges and with the same truncation of cavity as in S_m, and $E_{(all)}^{sol}$ is the total energy of a pure charge system with full background point charges and without truncation of cavity [see Fig. S1 (IV)]. Thus, in the GEBF-SMD approach, only the calculations on three types of systems (II, III, and IV) are needed, and the SMD calculation on the entire system is not required.

The total energy of the mth embedded subsystem (S_m) with truncated cavity is expressed as follows,

$$E_m^{sol} = E_m^a + E_m^b + E_m^{a,b} + \frac{1}{2} E_m^{a,c} + \frac{1}{2} E_m^{b,c}$$

(S2)

Here the E_m^a and E_m^b are the energies of atoms (a) and background charges (b), respectively, as,

$$E_m^a = \langle \Psi_m | H_m^{(0)} | \Psi_m \rangle$$

(S3)

$$E_m^b = \sum_{A \in S_m} \sum_{B > A \in S_m} \frac{Q_A Q_B}{|r_A - r_B|}$$

(S4)

The $E_m^{a,b}$ is the interaction energy between the atoms and background charges,

$$E_m^{a,b} = \left\langle \Psi_m \left| \sum_{A \in S_m} \frac{-Q_A}{|r - r_A|} \right| \Psi_m \right\rangle + \sum_{k \in S_m} \sum_{A \in S_m} \frac{Z_k Q_A}{|r_k - r_A|}$$

(S5)

The $E^{a,c}_m$ is the interaction energy between the atoms and surface or cavity charges (c),

$$E^{a,c}_m = \left\langle \Psi_m \left| \sum_{s \in S_m} \frac{-q_s}{|r - r_s|} \right| \Psi_m \right\rangle + \sum_{k \in S_m} \sum_{s \in S_m} \frac{Z_k q_s}{|r_k - r_s|}$$

(S6)

The $E^{b,c}_m$ is the interaction energy between the background charges and surface or cavity charges (c),

$$E^{b,c}_m = \sum_{A \in S_m} \sum_{s \in S_m} \frac{Q_A q_s}{|r_A - r_s|}$$

(S7)

It should be noted that in Eq. S1, the energy of surface or cavity charges (c), E^c_m, is not included, and only halves of $E^{a,c}_m$ and $E^{b,c}_m$ are incorporated in the SMD EP energy of the subsystem S_m.

By replacing the atoms in S_m with the corresponding point charges, we then obtain a pure charge system $S_{(m)}$ with full background point charges and with the same truncation of cavity as in S_m. The total energy of $S_{(m)}$ can be represented as,

$$E^{sol}_{(m)} = \sum_{A \in all} \sum_{B > A \in all} \frac{Q_A Q_B}{|r_A - r_B|} + \frac{1}{2} \sum_{A \in all} \sum_{s \in S_m} \frac{Q_A q_s}{|r_A - r_s|}$$

(S8)

For a pure charge system with full background point charges and without truncation of cavity, its total energy is

$$E^{sol}_{(all)} = \sum_{A \in all} \sum_{B > A \in all} \frac{Q_A Q_B}{|r_A - r_B|} + \frac{1}{2} \sum_{A \in all} \sum_{s \in all} \frac{Q_A q_s}{|r_A - r_s|}$$

(S9)

Finally, by substituting Eqs. S8 and S9 into Eq. S1 , we can obtain the following expression as shown in the main text,

$$E_{EP}^{GEBF}$$

$$= \sum_m C_m \left(E_m^{sol} - \frac{1}{2} \sum_{A \in alls} \sum_{\in S_m} \frac{Q_A q_s}{|r_A - r_s|} \right) + \frac{1}{2} \sum_{A \in alls} \sum_{\in all} \frac{Q_A q_s}{|r_A - r_s|} - \left(\sum_m C_m - 1 \right) \sum_{A \in}$$

(S10)

2. The GEBF-SMD energies of 2AJM in solution with different NPA charges

Table S1. The GEBF-SMD energies (in a.u.) of 2AJM in aqueous solution at the M06-2X/6-31G(d,p) level with the NPA charges obtained from three different types of subsystems.

subsystem	SCRF energy
naked	-10868.994338
embedded in background charges	-10868.994383
embedded in background and surface charges	-10868.994313