Communications in Information and Systems
Volume 19, Number 4, 405–423, 2019

Reproducing ensemble averaged electrostatics with Super-Gaussian-based smooth dielectric function: application to electrostatic component of binding energy of protein complexes

Shailesh K. Panday, Mihiri H. B. Shashikala, Arghya Chakravorty, Shan Zhao, and Emil Alexov*

Proteins constantly sample various conformations as they carry their biological function, including interacting with their partners, and this should be taken into account in any numerical protocol aiming at computing their thermodynamic properties. Here we report an application of previously reported Super-Gaussian-based smooth dielectric function (J Math Biol., 2019 Jul; **79**(2):631–672) to reproduce ensemble averaged electrostatics. This is an important achievement, since it dramatically reduces the computational demand for MM/PBSA applications and bypasses the necessity of long molecular dynamics simulations. Instead, a single frame, typically energy-minimized structure, in conjunction of Super-Gaussian-based smooth dielectric function, as implemented in DelPhi, can deliver ensemble averaged quantities. The approach is tested against ensemble averaged electrostatic component of binding free energy of protein-protein complexes. It is demonstrated that Super-Gaussian-based smooth dielectric function reproduces ensemble averaged quantities, resulting in correlation coefficients of about 0.8 and slope of the fitting line of 1.0.

1. Introduction

Electrostatics plays essential role in various processes in molecular biology [1, 2]. Being long-range, the electrostatics in some cases dominates other energies and effects, and becomes the most important contributor e.g. for computing pH-dependent phenomena [3, 4], pK_a predictions [5, 6], modeling electrostatic contribution to binding [7–9], folding [10, 11] and solvation [12–15]. The role of electrostatics in molecular biology is manifested by

*Corresponding author. ORCID: 0000-0001-5346-0156.

two energy terms: the polar solvation energy and screened Coulombic interaction energy. These energies can be computed separately or simultaneously depending on the protocol and software applied. However, the main distinction between different approaches for modeling electrostatics is the level of details for describing solute and solvent phases. Explicit models consider the solute and water at atomistic level of details, while implicit models treat the solute and solvent as continuum media with appropriate dielectric properties [13, 14]. The implicit models are much more computationally efficient [16], and at the same time in many cases the results delivered by implicit models are very similar to that from the explicit models [17–23]. Even more, the implicit solvent models were demonstrated to overcome the sensitivity issue concerning free energy calculations with different explicit water models [14]. As mentioned above, electrostatic energy is an essential component of the total energy of macromolecules and their assemblages. Frequent task in molecular biophysics is to compute folding or binding free energies [24]. Among various approaches to model these energies, the molecular mechanics (MM) Poisson-Boltzmann (PB) surface area (MM/PBSA) method is a trade-off of speed and accuracy. Typically, one generates representative structures (snap shots) via molecular dynamics (MD) simulations and then each structure is subjected to energy calculations. Thus, the MM component is taken from the corresponding force field used in MD simulations, the polar solvation energy is calculated with implicit model (PB) [25] or Generalized Born (GB) [26], and surface area is used to estimate the non-polar solvation energy. However, generating the ensemble of snapshots may be quite computationally expensive (long MD simulations) and could result in hundreds or thousand structures, which in turn should be subjected to polar and non-polar energies calculations. Thus, the traditional, trajectory-based MM/PBSA method cannot be applied on genome-scale investigations. The alternative is to use end-points approach, such that one takes into consideration only the starting and ending structures of the process being modeled. Thus, for modeling folding free energy one considers a representative structure of folded and a representative structure of unfolded states; for modeling binding free energy, one uses representative structure of bound molecules and representatives of unbound molecules. Such a protocol is very computationally efficient and can be used for large-scale modeling, however, it is questionable if the calculated energies will reproduce ensemble averaged quantities. Recently we suggested that Gaussian-based approach, typically used to define molecular surface [27–29], can be applied to deliver a smooth dielectric function spanning over the solute and solvent [15, 30]. The motivation for such an approach stems from the understanding that biological macromolecules are

not rigid bodies and they constantly sample different conformations [31]. The Gaussian-based smooth dielectric function is intended to model dielectric properties of the corresponding macromolecules and surrounding water layers to reflect their physical nature: namely the macromolecules are inhomogeneous objects, with cavities inside and surrounding waters are not as flexible as the bulk waters. Thus, it is unrealistic to model macromolecules as objects with fixed dielectric constant, as well as to expect that the water phase close to macromolecule has the same dielectric constant as the bulk phase [32–34] (indeed, it was demonstrated that the water molecules which lie close to the solute and are in the cavities within the solute have different dielectric responses and dynamics than those in the bulk region [35, 36]). Instead, Gaussian-based smooth dielectric function describes macromolecules as inhomogeneous dielectric objects, assigning low dielectric value to tightly packed hydrophobic core, while higher dielectric value to molecular cavities and interface, typically composed of polar and charged groups [15, 30, 34]. Furthermore, the water phase close to the macromolecule is assigned lower dielectric value compared with the bulk water to reflect decrease in water's ability to re-orient [15, 30, 34]. Other researchers have also emphasized on smooth transition between protein and the water phase and role of geometry to predict protein flexibility [37, 38]. Currently we developed and implemented in DelPhi [39, 40] two Gaussian models, the Gaussian [30, 41] and super-Gaussian [42]. The super-Gaussian model in DelPhi simply incorporates the super-Gaussian function into the setup of Gaussian-based dielectric model, and the resulting dielectric function has modeling difference from that of super-Gaussian dielectric model [42]. While both DelPhi protocols have appealing physics, it is not clear if they result in improvement of the end-points model, i.e. to enable it to deliver energies similar to ensemble averaged energies. This question was partially addressed for the Gaussian-based model on a set of 74 proteins, where we demonstrated that it delivers polar solvation energies very close to the energies calculated with thermodynamic integration (TI) [33]. However, both models, Gaussian and super-Gaussian, were not tested to check if they can calculate ensemble averaged electrostatic component of binding free energy. In this work, we carry out an extensive investigation of the ability of the super-Gaussian-based smooth dielectric function to mimic the conformational flexibility in terms of delivering electrostatic energy from a single structure and test the ability of super-Gaussian dielectric function to deliver ensemble averaged electrostatic component of binding energy on a set of 15 protein-protein complexes. A detailed study of the model and the underlying physical principles reveals that the super-Gaussian-based method in conjunction with

an energy minimized structure of a protein (various minimization protocols were investigated) can reproduce its ensemble average electrostatic component of binding free energy. The outcome of this work provides a convenient option enabling computation of the average electrostatic energies which can circumvent the time-consuming ensemble-based calculations.

2. Materials and methods

In this work, we are presenting an application of super-Gaussian-based dielectric function to test its ability to deliver ensemble averaged electrostatic component of binding energy on a set of 15 protein-protein complexes.

Super-Gaussian model in DelPhi The super-Gaussian model implemented in DelPhi is a simple extension of Gaussian model [39, 40]. The super-Gaussian function utilized in [42] is adopted to define the density at the position $\vec{r}$ for the i^{th} atom

$$g_i(\vec{r}) = \exp\left[\frac{-|\vec{r} - \vec{r_i}|^{2m}}{(\sigma R_i)^{2m}}\right] \tag{1}$$

where $\vec{r_i}$ is the center of the i^{th} atom, R_i is the Van der Waals radius of the i^{th} atom and σ is the relative variance. If we take $m = 1$, this gives rise to the original Gaussian model, while in super-Gaussian model, $m = 2$ is usually taken. Based on the density for each atom, the total density function for the atoms and overlapped area covered by multiple atoms is given by

$$g(\vec{r}) = 1 - \prod_{i=1}^{N_m}[1 - g_i(\vec{r})] \tag{2}$$

where N_m stands for the total number of atoms. The total density function $g(\vec{r})$ contains cross terms such as $g_i g_j$ to account for the density of the overlap region due to the i^{th} and j^{th} atoms. In this way, the density at an overlap region has a higher value than that generated by a single atom, but is still bounded in $[0, 1]$, because the range of $g(\vec{r})$ is $[0, 1]$. With such a total density, the inhomogeneous dielectric function for the Poisson-Boltzmann (PB) equation is defined as a weighted convex combination

$$\epsilon(\vec{r}) = g(\vec{r})\epsilon_{in} + (1 - g(\vec{r}))\epsilon_{ex} \tag{3}$$

where ϵ_{in} and ϵ_{ex} are the dielectric constants in the macromolecule and water respectively. Note that this dielectric function can be simply implemented in

DelPhi, but has some differences from that of the super-Gaussian dielectric model [42].

Dataset selection The dataset was constructed by requiring the entries to have experimental structures with resolution better than (≤ 2.0 Å) and to have experimentally measured binding free energies (for a purpose of another study). The cases were selected from PDBbind [43] database on July 12, 2019. The corresponding crystal structures for the chosen protein-protein binding cases were obtained from RCSB Protein Data Bank [44]. The dataset was refined to include only cases of dimers, i.e. cases where exactly two monomers are involved in complex formation. Similarly, the completeness of structure is also important, so those cases for which structures has missing regions in it were also discarded. Furthermore, those cases in which any of the partner is smaller than 30 amino-acids would be more like a peptide rather than a protein, therefore, such cases were also discarded. To avoid artifacts caused by ions, structures having ions at the interface or having ions of higher valency were removed. Whenever, the same pair of proteins appeared more than once with variant of any of the partners, only one pair (preferably wild type case if present) was kept to avoid redundancy in the dataset. The final dataset comprises of 15 protein-protein structures (PDB ids: 1AY7, 1BRS, 1FLE, 1J7D, 1KXV, 1MZW, 1OP9, 1XG2, 2G2U, 2I25, 2OOB, 3N4I, 3ZU7, 4LYL and 4ML7). These 15 cases are prepared for the further energy minimization and MD simulation required for the computation and comparative analysis of ensemble averaged electrostatic component of binding affinity ($\langle \Delta\Delta G^{bind}_{elec} \rangle$) and electrostatic component of binding energy from single structure $\Delta\Delta G^{bind}_{elec}(X, M)$ using different dielectric models. Where X represents one of the following single structures (GM: GB minimized, VM: vacuum minimized, EM: explicit TIP3P water minimized and Xtal: crystal structure), and M represents the dielectric model e.g. traditional two-dielectric (2-dielectric) or super-Gaussian-based (SuGauss) smooth dielectric function.

Preparation and energy minimization The missing hydrogen atoms in the structure were added using the reduce program available in AMBER suit of programs [45] to the monomers and the complex, keeping the protonation of titratable residues as at neutral pH, and all the histidin residues are kept singly protonated. The structures were prepared using the leap program available in AMBER [45]. The Amber ff14SB [46] force field is used for the proteins. Firstly, gas phase structures of the monomers and the complex were prepared and saved. Later, the gas phase systems (protein-monomers/complex) were solvated into a box of TIP3P [47] explicit waters,

where the edges of water box are at least 13 Å away from all the atoms of the protein. Appropriate number of monovalent counter-ions (Na^+ or Cl^-) were added to the box using addIons2 program to neutralize the total charge in the solvation box containing the protein system.

Four different representative structures were generated. Thus, the structure of the complex prepared in gas-phase represents the crystal structure and it will be denoted by $Xtal$ hereafter. The $Xtal$ structure is minimized in vacuum in two stages: (a) the structure is minimized for 10,000 steps consisting 8,000 steps of steepest descent (SD) followed by 2,000 steps of conjugate gradient (CG) algorithm with a restraint of 10 kcal.mol^{-1}.Å^{-2} on the non-hydrogen atoms. (b) again the final structure obtained from previous stage, 10,000 steps (8,000 SD and 2,000 CG) energy minimization is performed without restraint. During minimization the external dielectric constants was kept 80. The final structure obtained after the vacuum minimization will be referred by VM hereafter.

The $Xtal$ structure is also minimized using a implicit water Generalized Born (GB) model referred as OBC2 and given by Onufriev et al [48], in two stages as in case of vacuum minimization. The final structure obtained after the GB minimization will be referred as GM hereafter.

The $Xtal$ structure minimization in explicit water (TIP3P [47]) is achieved by first minimizing it for 10,000 steps (9,500 SD and 500 CG) with restraint of 50 kcal/mol on non-hydrogen atoms, followed by a heating cycle in which the system is heated from 0 K to 300 K in 6 stages each raising temperature by 50 K during 10 pico-second (ps) MD, after heating the system undergoes another round of minimization for 10,000 steps (8,000 SD and 2,000 CG) with very weak 0.05 kcal/mol restraint on the non-hydrogen atoms in the system. The final structure obtained this way is termed EM hereafter.

Equilibration and MD simulation The system is equilibrated with decreasing restrains 1.0, 0.5, 0.1, 0.04, 0.01 and 0.0 kcal.mol^{-1}.Å^{-2} during 100, 100, 100, 400, 4000 ps and 10 ns MD in constant temperature-pressure condition (NPT), the temperature is regulated to a target temperature 300 K using a langevin dynamics thermostat with collision frequency ($\gamma = 2.0$ ps^{-1}). The pressure is regulated to target pressure 1 bar using Berendsen barostat [49] with isotropic position scaling and relaxasation time 1.0 ps. Bonds lengths involving hydrogen atoms are constrained using SHAKE [50] to allow 2 fs timestep in MD simulation. All the simulations are done under Periodic Boundary Condition, the electrostatics calculations are done with cutoff 9.0 Å and utilize Particle Mesh Ewald Summation [51].

The 50 ns production simulation is also done in NPT condition with same set of parameters, coordinates are saved every 10 ps, resulting 5,000 snapshots every simulation trajectory. These simulations are performed using GPU enabled code pmemd.cuda [52] available in Amber suit of programs [45] on Palmetto cluster of Clemson University, SC. The same protocol is used for minimization and simulation of all the 15 protein-protein binding cases, where each case has 3 set of simulations, one each of both monomers and one for complex. In summary, 45 such 50 ns simulations are carried in this case study. The electrostatic component of binding energy ($\Delta\Delta G^{bind}_{elec} = \Delta\Delta G^{solv}_{polar} + \Delta\Delta G_{coulomb}$), here it included polar component of solvation and Coulombic interactions. If we label monomers and complex with A, B and AB respectively, then ensemble averaged electrostatic component of binding energy i.e. $\langle\Delta\Delta G^{bind}_{elec}\rangle$ is computed using Equation 4.

$$\langle\Delta\Delta G^{bind}_{elec}\rangle = \langle\Delta G^{bind}_{elec}(AB)\rangle - \langle\Delta G^{bind}_{elec}(A)\rangle - \langle\Delta G^{bind}_{elec}(B)\rangle \tag{4}$$

where $\langle\Delta G^{bind}_{elec}(A)\rangle$, $\langle\Delta G^{bind}_{elec}(B)\rangle$ and $\langle\Delta G^{bind}_{elec}(AB)\rangle$ are ensemble averaged electrostatic component for A, B and AB respectively, which are computed for 5,000 snapshots from the corresponding MD simulation trajectory.

Computation of $\boldsymbol{\Delta\Delta G^{bind}_{elec}(X, M)}$ using a Super-Gaussian-based model The protocol of computing electrostatic component of binding requires grid center and number of grids to be kept same for the complex and monomers DelPhi runs. For the complex run the scale and perfil are set 2.0 grids/Å and 70 respectively. The appropriate number of the grids and the center of grid box is extracted from the DelPhi run for complex and used for DelPhi runs of monomers. The total grid energy obtained for complex from the DelPhi run and the difference of sum of monomer grid energies from complex grid energy yields the desired electrostatic component of the binding energy. The same process is repeated for every set of varying reference-dielectric (ϵ_{ref}) and Gaussian/Super-Gaussian parameter σ. The Gaussian-multiplier (gm) is kept at 2.

3. Results and discussion

Here we present the results about electrostatic component of the binding energy of protein-protein complexes. The ensemble average electrostatic component of protein-protein binding free energy $\langle\Delta\Delta G^{bind}_{elec}\rangle$ is computed using the 5,000 snapshots obtained from 50 ns MD simulation for each of the 15 protein-protein binding cases as discussed in Section 2. These results are then compared with $\Delta\Delta G^{bind}_{elec}(X, 2-dielectric)$ (where X is used to indicate

the single structure used: crystallographic or energy minimized in 3 different ways) to check if traditional two-dielectric model can reproduce the ensemble averaged values. For this purpose, the internal dielectric constant ϵ_{in} used in the two-dielectric model is varied from 1 to 10 to test and choose an optimal internal dielectric value. A linear regression of $\Delta\Delta G^{bind}_{elec}(X, 2-dielectric)$ against $\langle\Delta\Delta G^{bind}_{elec}\rangle$ is done for each ϵ_{in}, the correlation coefficient, slope and y-intercept of regression line are obtained for each of the four single structures and shown in the Figure 1.

Among the four single structures used with two-dielectric model in combination with varying ϵ_{in} from 1 to 10, GB minimized (GM) and crystal structure (Xtal) achieved highest Pearson correlation coefficients (PCC) and regression line slope (m) close to 1, the y-intercept (c) in these cases range between 0 to 60 kcal/mol, as shown in Figure 1. The best-fit case linear regression parameters values using GM structure are: R = 0.65, m = 1.08 and c = 24.67 kcal/mol at $\epsilon_{in} = 1$. However, using $Xtal$ structure with the same $\epsilon_{in} = 1$ results in slope m = 1.65 which is considerably away from desired slope of 1.0. The fit is significantly improved if one uses $\epsilon_{in} = 2$ and the corresponding fitting parameters are PCC = 0.63, m = 0.88 and c = 27.72 kcal/mol. Further increasing of ϵ_{in} leads to decrease in slope of regression line implying that the ability of reproducing $\langle\Delta\Delta G^{bind}_{elec}\rangle$ using two-dielectric model over single structure $\Delta\Delta G^{bind}_{elec}(X, 2-dielectric)$ is deteriorating. Briefly, the best possible regression values for this investigation came using Generalized Born minimized structure at $\epsilon_{in} = 1$, and these fit parameter values are PCC = 0.65, m = 1.08 and c = 24.67 kcal/mol. Next step is to investigate the capability of the super-Gaussian-based dielectric function to reproduce the ensemble averaged $\langle\Delta\Delta G^{bind}_{elec}\rangle$ from single structure as attempted in this case (results are shown in Supplementary Information Figures S1–4 (http://intlpress.com/site/pub/files/_supp/cis/2019/0019/0004/CIS-2019-0019-0004-s003.pdf)). Here we present results with $\epsilon_{ref} = 2.0$ to avoid complexity of the presentation.

The linear regression fit parameter values of $\Delta\Delta G^{bind}_{elec}(GM, SuGauss)$ against $\langle\Delta\Delta G^{bind}_{elec}\rangle$ upon varying Gaussian parameter σ from 0.8 to 1.4 in steps of 0.1 for every ϵ_{ref} from 1 to 10 are studied. The first set of results are done with GM energy minimized structure using super-Gaussian-based dielectric (SuGauss) function. The near optimal fit occurs at ϵ_{ref} 1 and 2. The optimal fit parameter values are PCC = 0.75, m = 0.98 and c = 25.95 kcal/mol at $\sigma = 1.0$ and $\epsilon_{ref} = 2$. The correlation coefficients does not improve with increasing internal dielectric constant but the slope of the fit line goes up to 2 and higher at Gaussian parameter value $\sigma = 1.0$. In comparison to traditional two-dielectric model (see results for two-dielectric

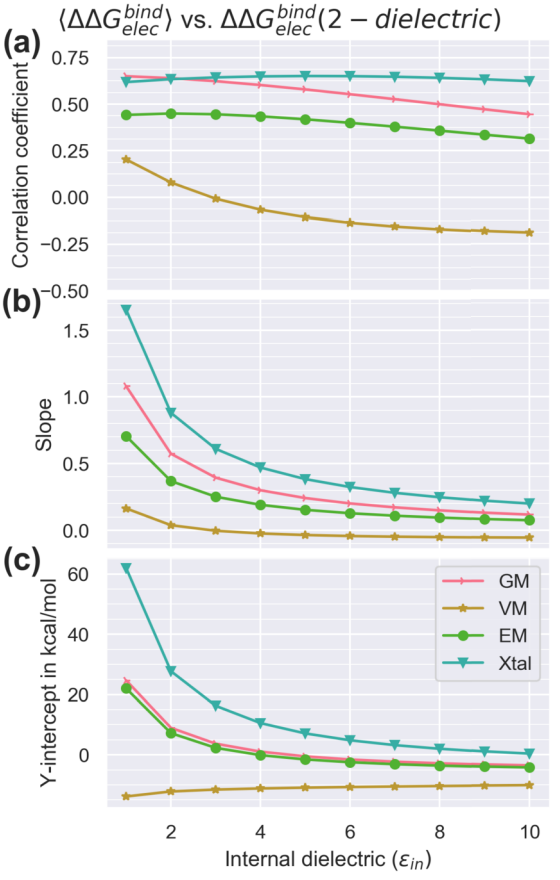

Figure 1: The fit parameter values obtained from linear regression of ensemble averaged electrostatic component of binding energy $\langle \Delta\Delta G^{bind}_{elec} \rangle$ against electrostatic component of binding energy using traditional two-dielectric over single structure $\Delta\Delta G^{bind}_{elec}$. GM: energy minimized in Generalized Born implicit solvent, VM: energy minimized in vacuum, EM: energy minimized in TIP3P explicit water and Xtal: Crystal structure correspond to treatment applied on the single structure are shown. (a) correlation coefficient, (b) slope of fit, and (c) y-intercept in kcal/mol. A data set of 15 protein-protein complexes as earlier described is used in this analysis.

model) here we obtained better regression line parameter values with super-Gaussian-based dielectric function when GB minimized single structure is used to predict the ensemble average electrostatic component of protein-protein binding energy.

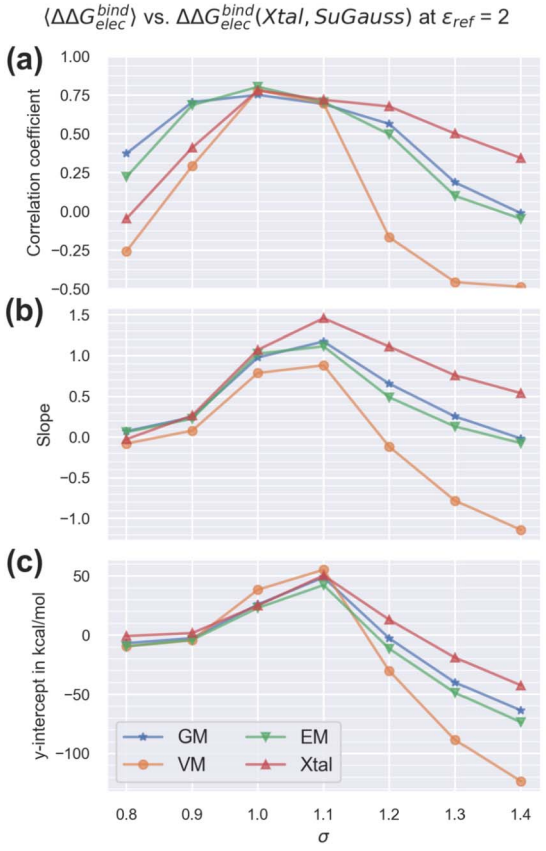

Figure 2: The fit parameter values obtained from linear regression of ensemble averaged electrostatic component of binding energy $\langle \Delta\Delta G^{bind}_{elec} \rangle$ against electrostatic component of binding energy using super-Gaussian dielectric function at various Gaussian parameter σ and reference dielectric $\epsilon_{ref} = 2$ over single structure, GM: energy minimized in Generalized Born implicit solvent, VM: energy minimized in vacuum, EM: energy minimized in TIP3P explicit water and Xtal: Crystal structure correspond to treatment applied on the single structure are shown. (a) correlation coefficient, (b) slope of fit, and (c) y-intercept in kcal/mol. A data set of 15 protein-protein complexes as earlier described is used in this analysis.

The trend observed for prediction of ensemble average electrostatic component of protein-protein binding free energy using super-Gaussian-based dielectric function using single energy minimized with GB model persist in case of vacuum minimized single structures as shown in Figure 2. However,

the optimal regression parameter values are PCC = 0.78, m = 0.79 and c = 38.38 kcal/mol at Gaussian parameter $\sigma = 1.0$ and $\epsilon_{ref} = 2$. In this case we obtain a slightly better correlation coefficient in comparison to GB case, but the slope of the regression line decreased further and goes away from 1.

In the case where single structure energy minimized in explicit TIP3P waters the correlation coefficient and slope both further improve from the cases of GM and VM single structures at the same Gaussian parameter value $\sigma = 1.0$ and internal dielectric $\epsilon_{ref} = 2$, the regression line parameters are shown in Figure 2, where Gaussian parameter σ is varied from 0.8 to 1.4 in steps of 0.1 at four different internal dielectric values 1, 2, 4 and 8. In this case the optimal regression line fit parameter values are PCC = 0.80, m = 1.03 and c = 23.06 kcal/mol. Here we observe significant improvement (PCC reaches 0.80 from 0.65, m moves close to 1, 1.03 from 1.08) in capability to reproducing ensemble averaged electrostatic component of binding free energy from single structure (energy minimized in explicit TIP3P water) of the super-Gaussian based dielectric function in comparison to traditional two-dielectric model.

The super-Gaussian-based dielectric function when used with single crystal structure also yields significantly better correlation coefficient in comparison to traditional two-dielectric model Figures 2 and 1. The linear regression parameter values in this case are PCC = 0.78, m = 1.07 and c = 25.07 kcal/mol at Gaussian parameter $\sigma = 1.0$ and $\epsilon_{ref} = 2$. Here, again we observe the regression line slope is higher than 1, in case of $\epsilon_{ref} = 1$, which gets close to 1 at $\epsilon_{ref} = 2$ and keeps on decreasing with increasing ϵ_{ref}. The correlation coefficient increases when Gaussian parameter σ is increasing from 0.8 to 1.0 but it changes the trend and starts decreasing when σ is further increased from 1. This implies the optimal *sigma* value lies in the close vicinity to 1.0, and optimal ϵ_{ref} is 2. Considering the best linear regression corresponding to $\sigma = 1.0$ with a coarse step size 0.1, we also investigated σ-dependence with fine step size 0.02. In this case σ was varied from 0.92 to 1.06 in steps of 0.02 and linear regression of $\Delta\Delta G^{bind}_{elec}(X, SuGauss)$ against $\langle \Delta\Delta G^{bind}_{elec} \rangle$ was performed and analyzed (see Supplementary Information Figure S5–8). However, the optimal σ and ϵ_{ref} are still 1.0 and 2 respectively. The similar trend is noticed in case of all the four single structure cases when used with super-Gaussian-based dielectric function for predicting the ensemble averaged electrostatic component of protein-protein binding energy.

The internal dielectric constant value $\epsilon_{ref} = 2$ is found to be optimal for all the four single structure cases. Therefore, to understand the

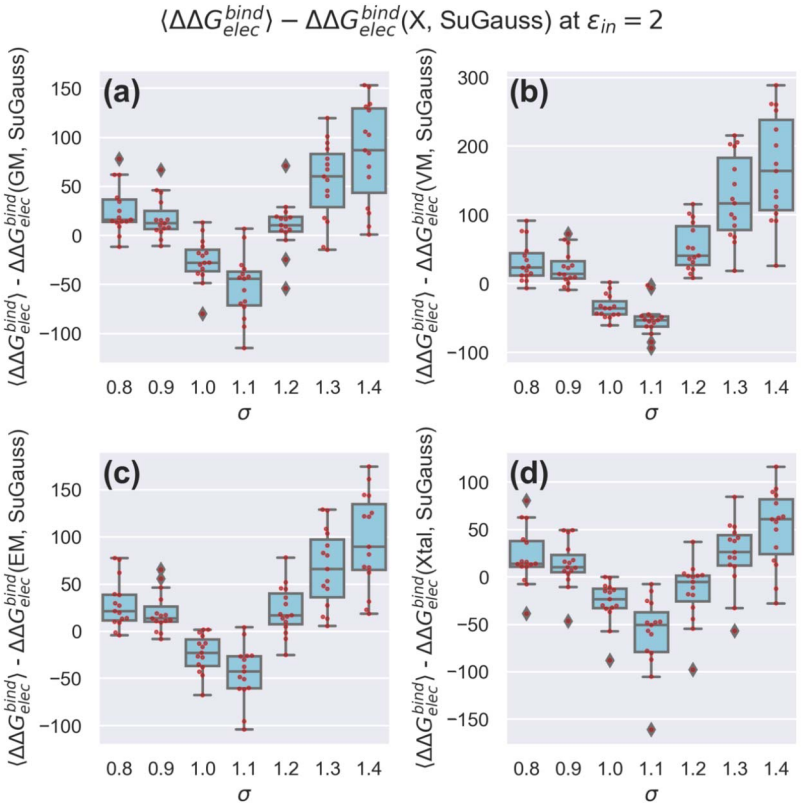

Figure 3: The boxplots of difference of electrostatic component of protein-protein binding energy obtained with super-Gaussian-based dielectric function using single structure $\Delta\Delta G^{bind}_{elec}(X, SuGauss)$ from ensemble averaged $\langle\Delta\Delta G^{bind}_{elec}\rangle$ at various σ. All the energies are in kcal/mol. Subplots show results obtained using single structure, (a) GB minimized, (b) vacuum minimized, (c) energy minimized in explicit TIP3P water and (d) crystal structure.

influence of varying σ on the ability of the super-Gaussian to reproduce ensemble average $\langle\Delta\Delta G^{bind}_{elec}\rangle$, we show the difference of $\langle\Delta\Delta G^{bind}_{elec}\rangle$ and $\Delta\Delta G^{bind}_{elec}(X, SuGauss)$ Figure 3. One can see that the best σ is 1.0, the range where the boxplots are narrowest and their mean values are close to zero. This is consistent with our previous Gaussian works where we indicated that optimal σ is 0.93 [15].

4. Conclusion

In this work we demonstrated that super-Gaussian model can reproduce ensemble averaged electrostatic component of the binding free energy. In all cases, the optimal reference dielectric value ϵ_{in} was found to be 2.0. The optimal values of the Gaussian parameter σ is approximately 1.0. We observe that the correlation coefficient (PCC) for electrostatic component of the binding free energy is much higher comparing with the PCC of traditional two-dielectric model. This indicates that super-Gaussian model is capable of predicting ensemble averaged electrostatic energies and can be used to mimic the effect of corresponding conformation changes.

References

[1] B. Honig and A. Nicholls, *Classical electrostatics in biology and chemistry*, Science, **268**(5214):1144–1149, 1995.

[2] A. Warshel and S. T. Russell, *Calculations of electrostatic interactions in biological systems and in solutions*, Quarterly Reviews of Biophysics, **17**(3):283–422, 1984.

[3] R. C. Mitra, Z. Zhang and E. Alexov, *In silico modeling of pH-optimum of protein–protein binding*, Proteins: Structure, Function, and Bioinformatics, **79**(3):925–936, 2011.

[4] J. Mongan, D. A. Case and J. A. McCammon, *Constant pH molecular dynamics in generalized Born implicit solvent*, Journal of Computational Chemistry, **25**(16):2038–2048, 2004.

[5] E. Alexov, E. L. Mehler, N. Baker, A. M. Baptista, Y. Huang, F. Milletti, J. Erik Nielsen, D. Farrell, T. Carstensen, M. H. Olsson et al., *Progress in the prediction of pKa values in proteins*, Proteins: Structure, Function, and Bioinformatics, **79**(12):3260–3275, 2011.

[6] L. Wang, L. Li and E. Alexov, *pKa predictions for proteins, RNA s, and DNA s with the Gaussian dielectric function using DelPhi pKa*, Proteins: Structure, Function, and Bioinformatics, **83**(12):2186–2197, 2015.

[7] M. Petukh, L. Dai and E. Alexov, *SAAMBE: webserver to predict the charge of binding free energy caused by amino acids mutations*, International Journal of Molecular Sciences, **17**(4):547, 2016.

[8] A. Chakavorty, L. Li and E. Alexov, *Electrostatic component of binding energy: interpreting predictions from poisson-boltzmann equation and modeling protocols*, Journal of Computational Chemistry, **37**(28):2495–2507, 2016.

[9] K. Talley, C. Ng, M. Shoppell, P. Kundrotas and E. Alexov, *On the electrostatic component of protein-protein binding free energy*, PMC Biophysics, **1**(1):2, 2008.

[10] I. Getov, M. Petukh and E. Alexov, *SAAFEC: predicting the effect of single point mutations on protein folding free energy using a knowledge-modified MM/PBSA approach*, International Journal of Molecular Sciences, **17**(4):512, 2016.

[11] R. Zhou, *Free energy landscape of protein folding in water: explicit vs. implicit solvent*, Proteins: Structure, Function, and Bioinformatics, **53**(2):148–161, 2003.

[12] R. Skyner, J. McDonagh, C. Groom, T. Van Mourik and J. Mitchell, *A review of methods for the calculation of solution free energies and the modelling of systems in solution*, Physical Chemistry Chemical Physics, **17**(9):6174–6191, 2015.

[13] J. Zhang, H. Zhang, T. Wu, Q. Wang and D. van der Spoel, *Comparison of Implicit and Explicit Solvent Models for the Calculation of Solvation Free Energy in Organic Solvents*, Journal of Chemical Theory and Computation, **13**(3):1034–1043, 2017.

[14] S. Izadi, B. Aguilar and A. V. Onufriev, *Protein–ligand electrostatic binding free energies from explicit and implicit solvation*, Journal of Chemical Theory and Computation, **11**(9):4450–4459, 2015.

[15] L. Li, C. Li and E. Alexov, *On the modeling of polar component of solvation energy using smooth Gaussian-based dielectric function*, Journal of Theoretical and Computational Chemistry, **13**(03):1440002, 2014.

[16] R. Anandakrishnan, A. Drozdetski, R. C. Walker and A. V. Onufriev, *Speed of conformational change: comparing explicit and implicit solvent molecular dynamics simulations*, Biophysical Journal, **108**(5):1153–1164, 2015.

[17] D. Shivakumar, Y. Deng and B. Roux, *Computations of absolute solvation free energies of small molecules using explicit and implicit solvent model*, Journal of Chemical Theory and Computation, **5**(4):919–930, 2009.

[18] A. Jean-Charles, A. Nicholls, K. Sharp, B. Honig, A. Tempczyk, T. F. Hendrickson and W. C. Still, *Electrostatic contributions to solvation energies: Comparison of free energy perturbation and continuum calculations*, Journal of the American Chemical Society, **113**(4):1454–1455, 1991.

[19] B. Jayaram, R. Fine, K. Sharp and B. Honig, *Free energy calculations of ion hydration: an analysis of the Born model in terms of microscopic simulations*, The Journal of Physical Chemistry, **93**(10):4320–4327, 1989.

[20] A. Mukhopadhyay, B. H. Aguilar, I. S. Tolokh and A. V. Onufriev, *Introducing charge hydration asymmetry into the generalized Born model*, Journal of Chemical Theory and Computation, **10**(4):1788–1794, 2014.

[21] A. V. Onufriev and B. Aguilar, *Accuracy of continuum electrostatic calculations based on three common dielectric boundary definitions*, Journal of Theoretical and Computational Chemistry, **13**(03):1440006, 2014. MR2998809

[22] A. Nicholls, D. L. Mobley, J. P. Guthrie, J. D. Chodera, C. I. Bayly, M. D. Cooper and V. S. Pande, *Predicting small-molecule solvation free energies: an informal blind test for computational chemistry*, Journal of Medicinal Chemistry, **51**(4):769–779, 2008.

[23] L. Y. Zhang, E. Gallicchio, R. A. Friesner and R. M. Levy, *Solvent models for protein–ligand binding: Comparison of implicit solvent Poisson and surface generalized Born models with explicit solvent simulations*, Journal of Computational Chemistry, **22**(6):591–607, 2001.

[24] A. Perez, J. A. Morrone, C. Simmerling and K. A. Dill, *Advances in free-energy-based simulations of protein folding and ligand binding*, Current Opinion in Structural Biology, **36**:25–31, 2016.

[25] J. Srinivasan, T. E. Cheatham, P. Cieplak, P. A. Kollman and D. A. Case, *Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices*, Journal of the American Chemical Society, **120**(37):9401–9409, 1998.

[26] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang et al., *Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models*, Accounts of Chemical Research, **33**(12):889–897, 2000.

[27] J. M. Word and A. Nicholls, *Application of the Gaussian dielectric boundary in Zap to the prediction of protein pKa values*, Proteins: Structure, Function, and Bioinformatics, **79**(12):3400–3409, 2011.

[28] J. Grant, B. Pickup, M. Sykes, C. Kitchen and A. Nicholls, *The Gaussian Generalized Born model: application to small molecules*, Physical Chemistry Chemical Physics, **9**(35):4913–4922, 2007.

[29] S. Decherchi, J. Colmenares, C. E. Catalano, M. Spagnuolo, E. Alexov and W. Rocchia, *Between algorithm and model: different molecular surface definitions for the Poisson-Boltzmann based electrostatic characterization of biomolecules in solution*, Communications in Computational Physics, **13**(1):61–89, 2013.

[30] L. Li, C. Li, Z. Zhang and E. Alexov, *On the dielectric "constant" of proteins: smooth dielectric function for macromolecular modeling and its implementation in DelPhi*, Journal of Chemical Theory and Computation, **9**(4):2126–2136, 2013.

[31] M. Levitt, *Protein conformation, dynamics, and folding by computer simulation*, Annual Review of Biophysics and Bioengineering, **11**(1):251–271, 1982.

[32] A. Chakravorty, Z. Jia, L. Li, S. Zhao and E. Alexov, *Reproducing the ensemble average polar solvation energy of a protein from a single structure: Gaussian-based smooth dielectric function for macromolecular modeling*, Journal of Chemical Theory and Computation, **14**(2):1020–1032, 2018.

[33] A. Chakravorty, Z. Jia, Y. Peng, N. Tajielyato, L. Wang and E. Alexov, *Gaussian-based smooth dielectric function: a surface-free approach for modeling macromolecular binding in solvents*, Frontiers in Molecular Biosciences, **5**:25, 2018.

[34] L. Li, L. Wang and E. Alexov, *On the energy components governing molecular recognition in the framework of continuum approaches*, Frontiers in Molecular Biosciences, **2**:5, 2015.

[35] W. Wang, O. Donini, C. M. Reyes and P. A. Kollman, *Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions*, Annual Review of Biophysics and Biomolecular Structure, **30**(1):211–243, 2001.

[36] S. K. Sinha, S. Chakraborty and S. Bandyopadhyay, *Thickness of the hydration layer of a protein from molecular dynamics simulation*, The Journal of Physical Chemistry B, **112**(27):8203–8209, 2008.

[37] D. D. Nguyen, K. Xia and G.-W. Wei, *Generalized flexibility-rigidity index*, The Journal of Chemical Physics, **144**(23):234106, 2016.

[38] L. Mu, K. Xia and G. Wei, *Geometric and electrostatic modeling using molecular rigidity functions*, Journal of Computational and Applied Mathematics, **313**:18–37, 2017. MR3573224

[39] C. Li, Z. Jia, A. Chakravorty, S. Pahari, Y. Peng, S. Basu, M. Koirala, S. K. Panday, M. Petukh, L. Li et al., *DelPhi Suite: New Developments and Review of Functionalities*, Journal of Computational Chemistry, 2019.

[40] L. Li, C. Li, S. Sarkar, J. Zhang, S. Witham, Z. Zhang, L. Wang, N. Smith, M. Petukh and E. Alexov, *DelPhi: a comprehensive suite for DelPhi software and associated resources*, BMC Biophysics, **5**(1):9, 2012.

[41] Z. Jia, L. Li, A. Chakravorty and E. Alexov, *Treating ion distribution with G aussian-based smooth dielectric function in DelPhi*, Journal of Computational Chemistry, **38**(22):1974–1979, 2017.

[42] T. Hazra, S. A. Ullah, S. Wang, E. Alexov and S. Zhao, *A super-Gaussian Poisson–Boltzmann model for electrostatic free energy calculation: smooth dielectric distribution for protein cavities and in both water and vacuum states*, Journal of Mathematical Biology, pages 1–42, 2019. MR3982707

[43] R. Wang, X. Fang, Y. Lu and S. Wang, *The PDBbind database: Collection of binding affinities for protein-ligand complexes with known three-dimensional structures*, Journal of Medicinal Chemistry, **47**(12):2977–2980, 2004.

[44] F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. F. Meyer Jr, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi and M. Tasumi, *The Protein Data Bank: a computer-based archival file for macromolecular structures*, Journal of Molecular Biology, **112**(3):535–542, 1977.

[45] D. A. Case, T. E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K. M. Merz Jr., A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *The Amber biomolecular simulation programs*, Journal of Computational Chemistry, **26**(16):1668–1688, 2005.

[46] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB*, Journal of Chemical Theory and Computation, **11**(8):3696–3713, 2015.

[47] D. J. Price and C. L. Brooks III, *A modified TIP3P water potential for simulation with Ewald summation*, The Journal of Chemical Physics, **121**(20):10096–10103, 2004.

[48] A. Onufriev, D. Bashford and D. A. Case, *Exploring protein native states and large-scale conformational changes with a modified generalized born model*, Proteins: Structure, Function, and Bioinformatics, **55**(2):383–394, 2004.

[49] H. J. Berendsen, J. v. Postma, W. F. van Gunsteren, A. DiNola and J. Haak, *Molecular dynamics with coupling to an external bath*, The Journal of Chemical Physics, **81**(8):3684–3690, 1984.

[50] J.-P. Ryckaert, G. Ciccotti and H. J. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes*, Journal of Computational Physics, **23**(3):327–341, 1977.

[51] T. Darden, D. York and L. Pedersen, *Particle mesh Ewald: An N. log (N) method for Ewald sums in large systems*, The Journal of Chemical Physics, **98**(12):10089–10092, 1993.

[52] R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand and R. C. Walker, *Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald*, Journal of Chemical Theory and Computation, **9**(9):3878–3888, 2013.

Shailesh K. Panday
Department of Physics and Astronomy
Clemson University
Clemson, SC 29631
USA
E-mail address: spanday@clemson.edu

Mihiri H. B. Shashikala
Department of Physics and Astronomy
Clemson University
Clemson, SC 29631
USA
E-mail address: mhewabo@g.clemson.edu

Arghya Chakravorty
Department of Physics and Astronomy
Clemson University
Clemson, SC 29631
USA
E-mail address: arghyac@g.clemson.edu

Shan Zhao
Department of Mathematics
University of Alabama
Tuscaloosa, AL, 35487
USA
E-mail address: szhao@ua.edu

Emil Alexov
Department of Physics and Astronomy
Clemson University
Clemson, SC 29631
USA
E-mail address: ealexov@g.clemson.edu

Received October 22, 2019