Finding Optimal Dielectric Boundary For Practical Continuum Solvent Calculations

Negin Forouzesh [1], Layne T. Watson [1,2,3] Alexey Onufriev [1,4]

[1]Department of Computer Science, Virginia Tech, Blacksburg, VA 24061, USA
[2]Department of Mathematics, Virginia Tech, Blacksburg, VA 24061, USA
[3]Department of Aerospace & Ocean Engineering, Virginia Tech, Blacksburg, VA 24061, USA
[4]Department of Physics, Virginia Tech, Blacksburg, VA 24061, USA

Abstract

Fast and accurate calculation of solvation free energies (ΔG) and binding free energies ($\Delta\Delta G$) is substantially important in many applications, e.g. Computer Aided Drug Design (CADD). Accurate determination of dielectric boundary (DB), plays a key role in such calculations. We run a massively parallel deterministic global optimization to find the optimal DB in terms of calculating ΔG and $\Delta\Delta G$.

Introduction

Solvent Models. Calculation of the solvation free energy is remarkably sophisticated and highly-dependant on the employed solvent model. There are two major types of solvation representation:

- **Explicit solvent model**: Arguably accurate but computationally slow.
- **Implicit solvent model**: Relatively less accurate but significantly fast.

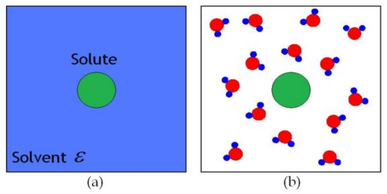


Figure 1: (a) Implicit solvent model (b) Explicit solvent model.

Solvation Free Energy. To estimate the total solvation free energy of a molecule, ΔG_{solv}, it is decomposed into the "electrostatic" and "non-electrostatic" parts:

$$\Delta G_{solv} = \Delta G_{el} + \Delta G_{nonel} \tag{1}$$

where ΔG_{nonel} is the free energy of solvating a molecule from which all charges have been removed. ΔG_{el} is the free energy if first all charges are removed in the vacuum, and then they are added back in the presence of a continuum solvent environment.

Binding Free Energy. The electrostatic component of binding free energy is computed by following the steps of this thermodynamics cycle:

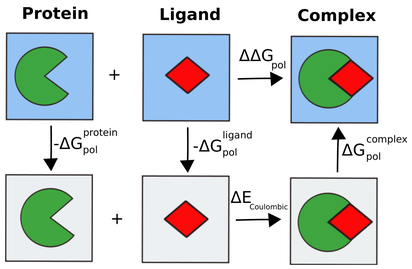


Figure 2: The thermodynamic cycle used for computing the binding free energy. Water environment is shown in blue, and vacuum is in white.

Dielectric Boundary. A key step in these continuum calculations is defining the solute/ solvent dielectric boundary (DB) – a region of space over which the dielectric constant $\epsilon(\mathbf{r})$ changes from the value characteristic of the molecular interior ($\epsilon = 1$) to that of the solvent, ($\epsilon = 80$) for water.

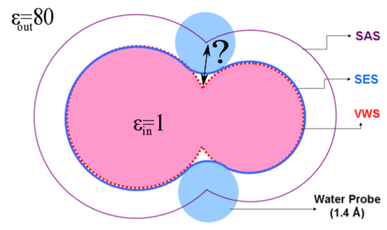


Figure 3: Molecular surface definitions. Dielectric boundary separates solvent/solute with two different dielectric constants (ϵ_{in} and ϵ_{out}).

Method

Implicit Solvent Details. GBNSR6 is an implementation of the GB model in which the effective Born radii are computed numerically, via the so-called "R6" integration, over the Lee-Richards molecular surface [1]. The analytical liniearized Poisson-Boltzmann (ALPB) model is used to approximate ΔG_{pol}, using the following formula:

$$\Delta G_{el} \approx -\frac{1}{2}\left(\frac{1}{\epsilon_{in}} - \frac{1}{\epsilon_{out}}\right)\frac{1}{1+\beta\alpha}\sum_{ij} q_i q_j \left(\frac{1}{f^{GB}} + \frac{\alpha\beta}{A}\right) \tag{2}$$

We employed the most widely used functional form of $f^{GB} = [r_{ij}^2 + R_i R_j exp(-r_{ij}^2/4R_iR_j)]^{1/2}$ where R_i is the so-called effective Born radius of atom i, and r_{ij} is the distance between atoms i and j. We set $\epsilon_{in} = 1$ and $\epsilon_{out} = 80$ in equation 2. The **effective Born radii** R were calculated by the following equation:

$$R_i^{-3} = \left(-\frac{1}{4\pi}\oint_{\partial V}\frac{\mathbf{r}-\mathbf{r_i}}{|\mathbf{r}-\mathbf{r_i}|^6}\cdot d\mathbf{S}\right) \tag{3}$$

where ∂V represents the molecular surface of the molecule, $d\mathbf{S}$ is the infinitesimal surface element vector, r_i is the position of atom i, and r represents the position of the infinitsimal surface element.

Molecular Surface Representation. The grid-based molecular surface used here is based on the "field-view" method [2] (Fig4 and Fig5).

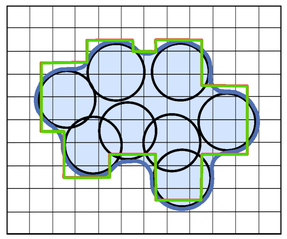


Figure 4: Finite-difference discretization of an abstract molecule.

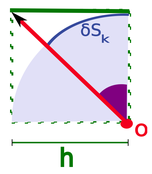


Figure 5: Spherical surface element δS_k (blue) and the corresponding square surface element (green).

DIviding-RECTangles (DIRECT) Global Search Algorithm.

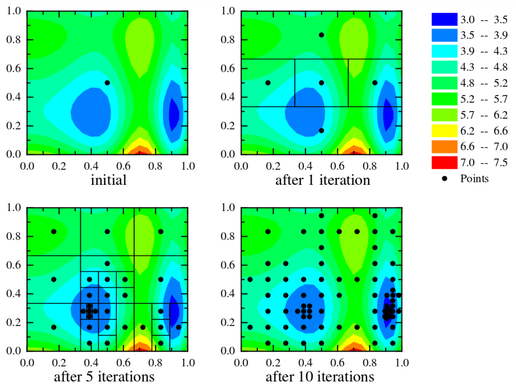


Figure 6: Function evaluations performed by DIRECT [3] after 0, 1, 5 and 10 iterations.

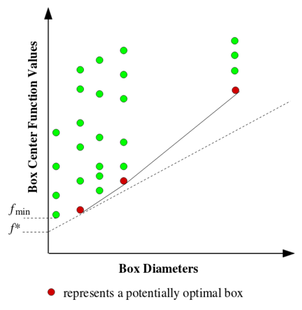


Figure 7: Global convergence property of DIRECT algorithm.

Results

Objective Function and Constraints. The objective function to be minimized is as follows:

$$RMS_{total} = 1.0 \times RMS_{\Delta\Delta G_{el}} + 0.3 \times RMS_{\Delta G_{el}} \tag{4}$$

The reference model is set to a well-known explicit solvent model, TIP3P [4]. Radial Distribution Function(RDF) is used to determine the feasible range for probe radius (ρ_w) and the following atomic radii (ρ_i): Carbon (C), Hydrogen (H), Nitrogen (N), Oxygen (O), Sulfur (S).

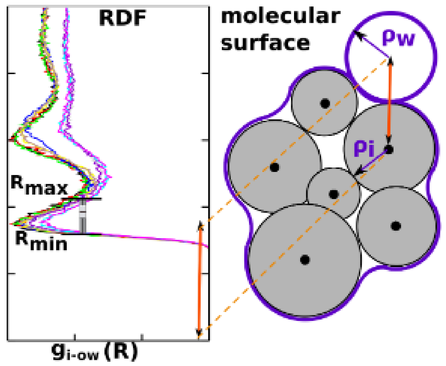


Figure 8: Optimal dielectric boundary (DB) based on the Radial Distribution Function (RDF) as the physical constraint.

Constraints based on the RDF:

$$\begin{aligned} 0.2 &\le r_W \le 1.6 \\ 2.2 &\le r_W + r_C \le 3.8 \\ 1.4 &\le r_W + r_H \le 3.0 \\ 2.2 &\le r_W + r_N \le 3.8 \\ 2.2 &\le r_W + r_O \le 3.8 \\ 2.9 &\le r_W + r_S \le 4.2 \end{aligned} \tag{5}$$

Preliminary Results. VTDIRECT95 [5] is a Fortran 95 implementation of DIRECT which is widely used in various applications.

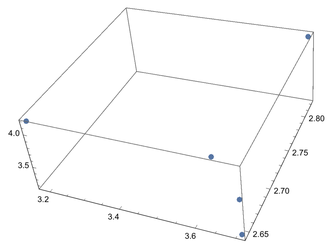


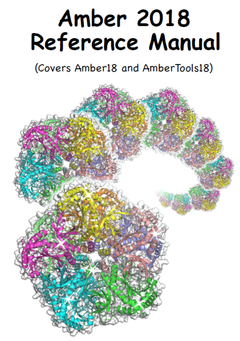
Figure 9: Five best optimals projected on a 3D box made of H, O and S.

Table 1: Five best box centers in the Exploration phase compared to BONDI radii. Last column shows the result of Exploitation around OPT1.

	REFERENCE	EXPLORATION					EXPLOITATION
ITEM	**BONDI**	OPT1	OPT2	OPT3	OPT4	OPT5	**OPT1**
r_W	**1.4**	1.37	1.37	1.52	1.37	1.37	**1.49**
r_C	**1.7**	1.48	1.46	1.77	1.69	1.46	**1.72**
r_H	**1.2**	1.54	1.54	1.45	1.54	1.54	**1.48**
r_N	**1.55**	2.28	2.34	2.19	1.81	2.34	**2.29**
r_O	**1.5**	1.28	1.28	1.30	1.28	1.28	**1.28**
r_S	**1.8**	2.83	2.33	2.65	2.81	1.80	**2.71**
$RMS_{\Delta\Delta G_{el}}$	**5.42**	3.69	3.87	4.02	4.06	4.00	**3.64**
$RMS_{\Delta G_{el}}$	**2.10**	2.25	2.27	2.00	1.94	2.27	**2.12**
RMS_{total}	**6.05**	4.37	4.55	4.62	4.64	4.68	**4.28**
$RMS_{\Delta\Delta G_{el}}$	**8.69**	6.59	6.67	6.61	6.37	6.65	**6.40**
$RMS_{\Delta G_{el}}$	**1.99**	2.46	2.52	2.19	2.02	2.52	**2.35**
RMS_{total}	**9.29**	7.33	7.42	7.27	6.97	7.41	**7.11**

Future Works

- **More comprehensive datasets.** Already, a set of 15 small protein-ligand complexes [6] for calculating $\Delta\Delta G_{pol}$, and a set of 90 small molecules [7] for calculating ΔG_{pol} are used. We propose to expand this datasets to have a more realistic optimization.
- **Different solvent models.** In this study we employed GBNSR6 for calculating the free energies, when TIP3P results were the reference. By testing different solvent models we can have a better understanding of the optimal dielectric boundary.
- **Implementation in AMBER.** Amber is a suite of biomolecular simulation programs. It is a package of molecular simulation programs which includes source code and demos. We propose to implement the final results in AMBER to make it accessible to all the users.



Conclusion

A novel systematic approach is proposed to constructing optimal, yet physically realistic dielectric boundaries for a range of continuum electrostatic theories. The approach differs from past efforts in several critical aspects:

- **Contribution 1.** Physical realism is enforced by experimentally observable atom-(water oxygen) RDFs.
- **Contribution 2.** The optimal parameters of the DB are well-suited for protein-ligand binding calculations through inclusion of protein-ligand energies and the corresponding molecular structures into the training set.

The result of this research will benefit computer scientists who work in Computer Aided Drug Design (CADD) research field and pharmaceutical industry.

References

[1] N. Forouzesh, S. Izadi, and A. V. Onufriev, "Grid-based surface generalized born model for calculation of electrostatic binding free energies," *Journal of Chemical Information and Modeling*, 2017.

[2] Q. Cai, X. Ye, J. Wang, and R. Luo, "On-the-fly numerical surface integration for finite-difference poisson–boltzmann methods," *Journal of chemical theory and computation*, vol. 7, no. 11, pp. 3608–3619, 2011.

[3] D. R. Jones, C. D. Perttunen, and B. E. Stuckman, "Lipschitzian optimization without the lipschitz constant," *Journal of optimization Theory and Applications*, vol. 79, no. 1, pp. 157–181, 1993.

[4] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *The Journal of chemical physics*, vol. 79, no. 2, pp. 926–935, 1983.

[5] J. He, L. T. Watson, and M. Sosonkina, "Algorithm 897: Vtdirect95: serial and parallel codes for the global optimization algorithm direct," *ACM Transactions on Mathematical Software (TOMS)*, vol. 36, no. 3, p. 17, 2009.

[6] 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*, vol. 11, no. 9, pp. 4450–4459, 2015.

[7] A. Mukhopadhyay, I. S. Tolokh, and A. V. Onufriev, "Accurate evaluation of charge asymmetry in aqueous solvation," *The Journal of Physical Chemistry B*, vol. 119, no. 20, pp. 6092–6100, 2015.

Contact Information

Email: neginf@vt.edu
Linkedin: linkedin.com/in/negin-forouzesh-726a395b