

# Hybrid Physical and Knowledge based Model for Blind Solvation Free Energy Prediction

Bao Wang, Zhixiong Zhao and Guo-wei Wei

Department of Mathematics, Michigan State University, East Lansing, MI, 48824



## INTRODUCTION AND MODELS

Solvation process is omnipresent in the chemical and biological systems which is mainly measured by the solvation free energy, effective and accurate solvation free energy prediction is of critical importance in understanding the solvation process. In the classical model, the solvation free energy  $\Delta G$  is decomposed into two parts, the polar part  $\Delta G^p$  which measures the electrostatics interaction, and the nonpolar part  $\Delta G^{np}$  which characterizes the hydrophobic interaction. In this work, we present the coupling of physical model and statistical model for the highly accurate solvation free energy prediction, in which the physical model is used for modeling of  $\Delta G^p$ , meanwhile the statistical model is employed for modeling  $\Delta G^{np}$ .

## PHYSICAL MODELS

The physical model adopted in this work for modeling of the electrostatics interaction is the Poisson Boltzmann (PB) model, which is formulated mathematically as:

$$-\nabla \cdot (\epsilon(\mathbf{r})\nabla\phi) = \rho_{total} + \sum_{i=1}^{N_c} Q_i e^{-Q_i\phi/k_B T}, \quad (1)$$

where  $\phi$  is the electrostatics potential field,  $\epsilon(\mathbf{r})$  is the permittivity function which takes  $\epsilon^+$  in the solvent domain while  $\epsilon^-$  in the solute domain,  $\rho_{total}$  used to describe the solute charge,  $\sum_{i=1}^{N_c} Q_i e^{-Q_i\phi/k_B T}$  used for the description of the solvent charge description in which,  $N_c$  is the number of ion species in the solvent,  $Q_i$  is the charge of the corresponding ion species,  $k_B$  are the Boltzmann constant, and  $T$  are the absolute temperature.

To make the PB model well posed in the computational sense, we enforce the Debye-Huckel boundary condition:

$$\phi(\mathbf{r}_i) = \int \frac{\rho_{total}}{\epsilon^+|\mathbf{r}-\mathbf{r}_i|} d\mathbf{r}, \quad (2)$$

where  $\mathbf{r}_i$  is the position of the given point on the boundary of the computational domain  $\partial\Omega$ . Across the interface  $\Gamma$  of the solute and solvent domain, the following interface conditions are enforced:

• the continuity of the electrostatics potential:

$$[\phi]_{\Gamma} = \phi^+(\mathbf{r}) - \phi^-(\mathbf{r}) = 0$$

• the continuity of the electrostatics flux:

$$[\epsilon\phi_{\mathbf{n}}] = \epsilon^+(\mathbf{r})\nabla\phi^+(\mathbf{r}) \cdot \mathbf{n} - \epsilon^-(\mathbf{r})\nabla\phi^-(\mathbf{r}) \cdot \mathbf{n} = 0,$$

where  $\mathbf{n} = (n_x, n_y, n_z)$  is the outer normal direction of the interface  $\Gamma$  which pointing from the solute domain to the solvent domain.

The solvent excluded surface (SES) generated by the ESES software is used for describing the solute molecule conformational structure.

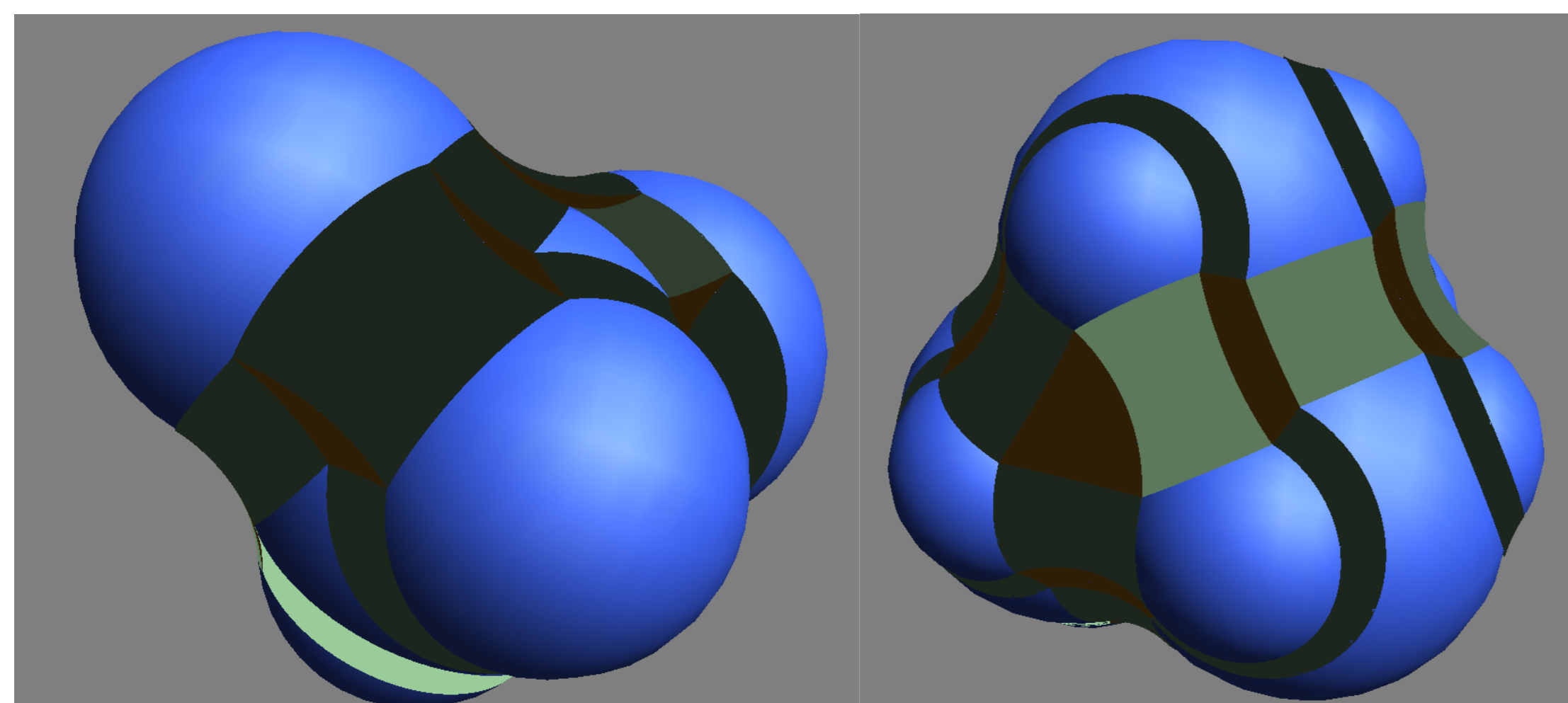


Figure 1. The SES of the solute molecule dichloromethane and neopentane.

Figure 1 depicts the SES surface for the solute molecules dichloromethane and neopentane, respectively.

The charge of the solute molecules can be considered in two approaches, namely, the semi-empirical charge from the Amber force field and from the Ab initio charge calculation.

## PHYSICAL MODELS

**Semi-empirical Charge:** In this work three types of semi-empirical charge from the Amber force field are employed for the charge parameterization of the solute molecules, namely, the AM1-BCC, Mulliken, and Gasteiger charges.

**Ab Initio Charge:** In our model, the Ab initio charge is obtained by using density function theory (DFT) from solving the Kohn Sham equation:

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + U_{eff}\right)\psi_j = E_j\psi_j, \quad (3)$$

where  $U_{eff}$  is the effective Kohn-Sham potential,  $\psi_j$  are the corresponding Kohn-Sham orbital. The charge density of the solute molecular can be obtained through taking the summation over all the Kohn Sham orbital, i.e.,  $\rho_{total} = \sum |\psi_j|^2$ .

To incorporate the solvent effects in the charge calculation, we add the reaction field energy to the effective Kohn-Sham potential of the KSDFT in the vacuum:

$$U_{eff} = U_{eff}^0 + \int \rho_{total}\phi_{RF},$$

where  $U_{eff}^0$  is the effective Kohn-Sham potential in vacuum,  $\phi_{RF}$  is the reaction field potential defined as the difference of the electrostatics potential in solvent and vacuum.

## STATISTICAL MODELS

In our work, the nonpolar solvation free energy  $\Delta G^{np}$  for the monofunctional group molecule is modeled as:

$$\Delta G^{np} = \sum_{i=1}^N \gamma_i Area_i + pVol, \quad (4)$$

where  $N$  is the total number of atom types in the solute molecules considered.  $\gamma_i$  denotes the atomic surface tension for the  $i$ th type atom,  $Area_i$  denotes the atomic area contributed from the  $i$ th type atom,  $p$  is the hydrodynamic pressure,  $Vol$  is the volume occupied by the solute molecule.

Our statistical model for the prediction of  $\Delta G^{np}$  can be split into two stages:

- **Stage 1.** Learning the atomic surface tension  $\gamma_i$  and hydrodynamic pressure  $p$  for each group of monofunctional group molecules in the training set based on the Tikhonov regularized least square regression.
- **Stage 2.** Scoring the contribution of each functional group influence on the nonpolar solvation free energy based on the convex optimization by using the polyfunctional group molecules in the training set.

For an arbitrary molecule in the testing set, we predict the corresponding  $\Delta G^{np}$  based on the regression parameters and scores of each functional group learned from the above statistical model. For a polyfunctional group molecule, we first apply the regression model for each functional group to get the corresponding predictions, second we ensemble each contributions based on the weights learned in the second stage.

## NUMERICAL METHOD

Our numerical method contains the following aspects:

- Generate the SES by the ESES software.
- Solve the PB model by the MIBPB software.
- Solve the Kohn-Sham DFT by the SIESTA software.
- Coupling the Kohn-Sham DFT and PB through the communication of the reaction field energy and solute charges, during the coupling the total charge conserving scheme is applied for the communication. And solve the coupling KSDFT and PB in a self consistent manner.
- Calculate the atomic surface area and volume of the solute molecule by the ESES software.
- Scoring the influence of each functional group through solving the constrained optimization problem through introducing the Lagrangian multiplier.

## RESULTS AND DISCUSSION

Figure 2 and 3 depicts the predicted and experimental solvation free energy for the SAMPL3 and SAMPL4 blind prediction test set. The RMS error for the SAMPL3 and SAMPL4 predictions are 0.77 and 1.03 kcal/mol, respectively. In this prediction, the Amber mbondi2 force field is used for the atomic radius parameterization. The Gasteiger charge is utilized for the charge assignment of the SAMPL3 test set, while the Ab initio charge calculation from coupling of MIBPB and SIESTA are employed for the SAMPL4 test set.

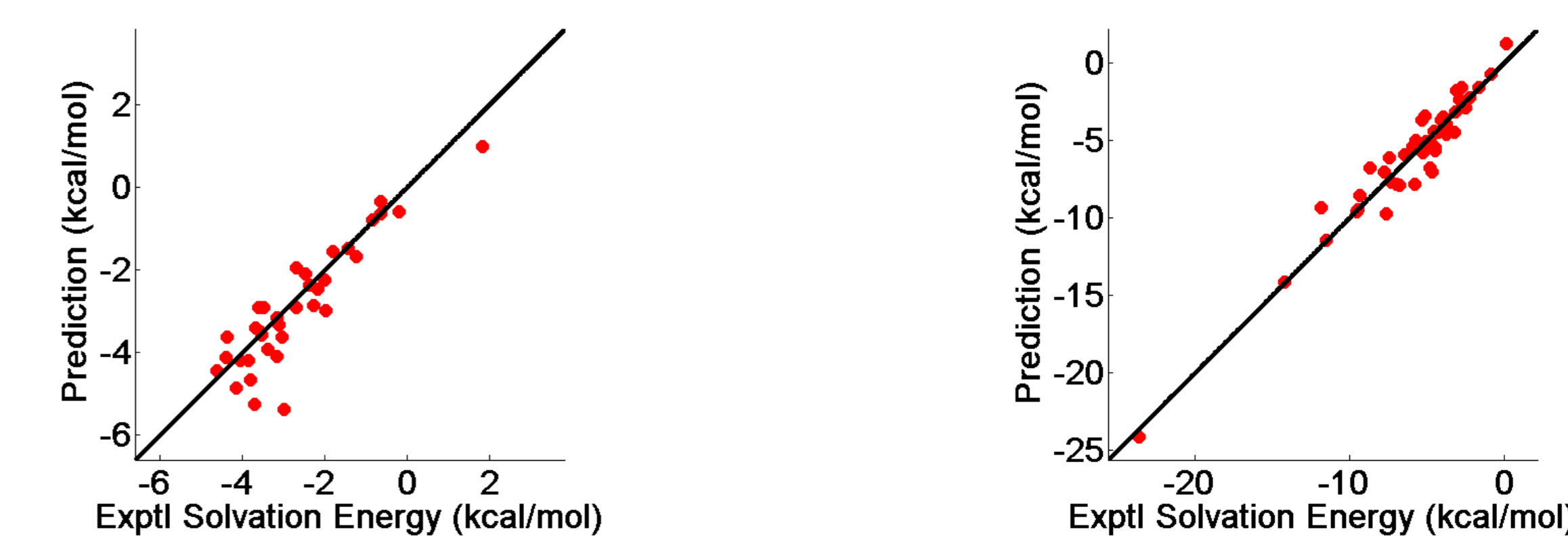


Figure 2. The comparison of predicted and experimental solvation free energy for SAMPL3 set (Left chart) and SAMPL4 set (Right chart). Their RMS errors are 0.77 kcal/mol and 1.03 kcal/mol, respectively.

Figure 3 demonstrates the influence of the force field parameters on the solvation free energy prediction. Here four different atomic radius and charges are compared for the solvation free energy prediction. The atomic radius considered are Amber 6, Amber bondi, Amber mbondi2, and ZAP9 force field. The charge considered ranging from semi-empirical to Ab-initio charge with the consideration of the solvent polarization to the charge distribution.

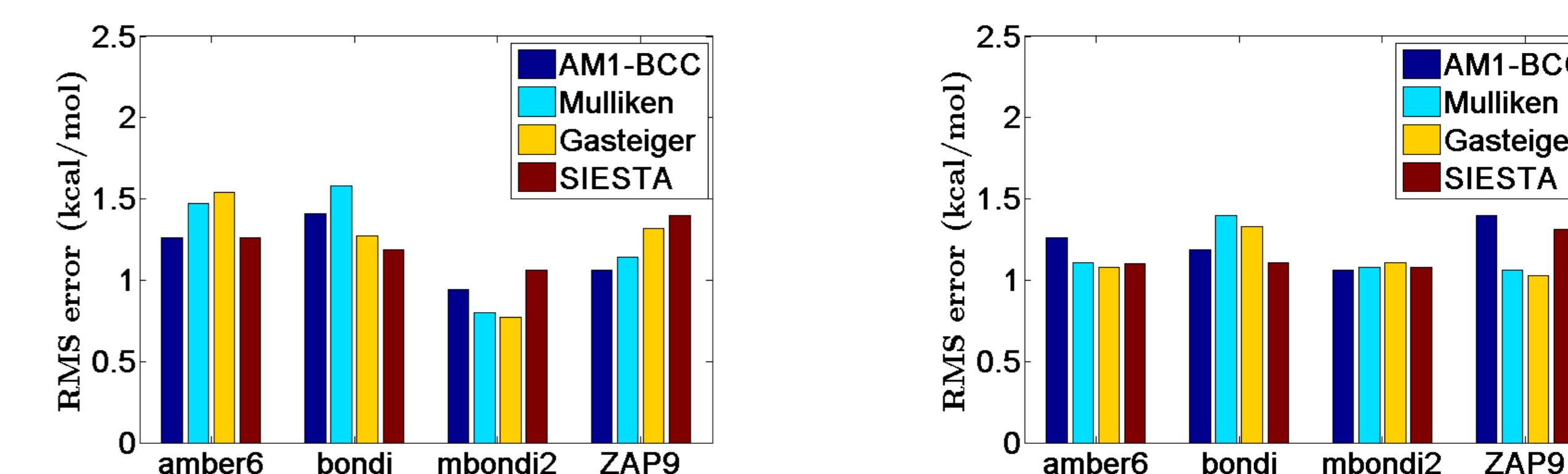


Figure 3. The influence of the atomic radius and charge force field on the solvation free energy prediction. Their influence on the SAMPL3 and SAMPL4 test set are illustrated in the left and right chart, respectively.

The results in fig. 3 indicates that the Amber mbondi2 atomic radii is the most stable one for the solvation free energy prediction. For the SAMPL3 set, the Gasteiger charge is optimal, while for the SAMPL4 test set the coupling of the MIBPB and SIESTA charge is the best one.

## REFERENCES

1. Bao Wang, Zhixiong Zhao, Guo-Wei Wei, Coupling of Physical and Statistical Model for Highly Accurate Solvation Prediction, preprint, 2015.
2. Bao Wang, Guo-Wei Wei, A coarse grid Poisson Boltzmann solver without loss of accuracy, preprint, 2015.
3. Beibei Liu, Bao Wang, Rundong Zhao, Yiyong Tong and Guo-Wei Wei, ESES: software for Eulerian solvent excluded surface, preprint, 2015.

## ACKNOWLEDGEMENT



This work was supported in part by NSF grants IIS-1302285 and DMS-1160352, NIH grant R01GM-090208 and MSU Center for Mathematical Molecular Biosciences Initiative.

