

# Parameter Optimization in Differential Geometry based Solvation Models

Bao Wang and Guo-wei Wei

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



## INTRODUCTION AND MODELS

Differential geometry based solvation models are a new class of implicit solvent models that are able to avoid unphysical solvent-solute boundary definitions and associated geometric singularities. The model dynamically coupled the polar and nonpolar solvation free energies, and the governing equations of the differential geometry based solvation models can be derived by the minimization of total solvation free energy functional.

In our model, the total solvation free energy is partitioned into two parts, the polar and nonpolar components:

$$G^{\text{Total}} = G^{\text{P}} + G^{\text{NP}}, \quad (1)$$

where  $G^{\text{P}}$  and  $G^{\text{NP}}$  are polar and nonpolar solvation energies, respectively. The polar energy represents the electrostatics part, and the nonpolar part counts the other solvent solute interactions, which is given by:

$$G^{\text{NP}} = \gamma \text{Area} + p \text{Vol} + \int_{\Omega_s} U d\mathbf{r}, \quad \mathbf{r} \in \mathbb{R}^3, \quad (2)$$

where, Area and Vol are the surface area and volume of the solute, respectively;  $\gamma$  is the surface tension;  $p$  is the hydrodynamic pressure;  $U$  denotes the solvent-solute non-electrostatic interactions; and  $\Omega_s$  is the solvent domain.

Based on the geometric measure theory, the nonpolar free energy can be formulated as:

$$G^{\text{NP}}[S] = \int [\gamma |\nabla S| + pS + (1-S)U] d\mathbf{r} \quad (3)$$

where  $0 \leq S \leq 1$  is a hypersurface or simply surface function that characterizes the solute domain and embeds the 2D surface in  $\mathbb{R}^3$

Assume that the aqueous environment has multiple species labeled by  $\alpha$ , and their interactions with each solute atom near the interface can be given by:

$$U = \sum_{\alpha} \rho_{\alpha} U_{\alpha} = \sum_{\alpha} \rho_{\alpha}(\mathbf{r}) \sum_j^{N_m} U_{\alpha j}(\mathbf{r}) \quad (4)$$

where  $\rho_{\alpha}(\mathbf{r})$  is the density of  $\alpha$ th solution component, which may be charged or uncharged,  $N_m$  is the number of atoms in the solute, and  $U_{\alpha j}$  is an interaction potential between the  $j$ th atom of the solute and the  $\alpha$ th component of the solvent. When water is used and there is free of other species,  $\rho_{\alpha}(\mathbf{r})$  is the water molecule density.

In the model, the polar solvation free energy functional can be expressed as:

$$G^{\text{P}} = \int \left\{ S \left[ -\frac{\epsilon_m}{2} |\nabla \Phi|^2 + \Phi \rho_m \right] + (1-S) \left[ -\frac{\epsilon_s}{2} |\nabla \Phi|^2 - k_B T \sum_{\alpha} \rho_{\alpha 0} \left( e^{-\frac{q_{\alpha} \Phi}{k_B T}} - 1 \right) \right] \right\} d\mathbf{r}, \quad (5)$$

where  $\epsilon_s$  and  $\epsilon_m$  are the dielectric constants of the solvent and solute, respectively. Here  $k_B$  is the Boltzmann constant,  $T$  is the temperature,  $\rho_{\alpha 0}$  denotes the reference bulk concentration of the  $\alpha$ th solvent species, and  $q_{\alpha}$  denotes the charge valence of the  $\alpha$ th solvent species, which is zero for an uncharged solvent component. We use  $\rho_m$  to represent the charge density of the solute. The charge density is often modeled by a point charge approximation

$$\rho_m = \sum_j^{N_m} Q_j \delta(\mathbf{r} - \mathbf{r}_j), \quad (6)$$

where  $Q_j$  denoting the partial charge of the  $j$ th atom in the solute.

Based on the variational principle, we construct the following generalized Laplace-Beltrami (GLB) equation:

$$\frac{\partial S}{\partial t} = |\nabla S| \left[ \nabla \cdot \left( \gamma \frac{\nabla S}{|\nabla S|} \right) + V \right], \quad (7)$$

where the potential driven term is given by

$$V = -p + U + \frac{\epsilon_m}{2} |\nabla \Phi|^2 - \Phi \rho_m - \frac{\epsilon_s}{2} |\nabla \Phi|^2 - k_B T \sum_{\alpha} \rho_{\alpha 0} \left( e^{-\frac{q_{\alpha} \Phi}{k_B T}} - 1 \right). \quad (8)$$

As in the nonpolar case, solving the generalized Laplace-Beltrami equation (7) generates the solvent-solute interface through the function  $S$ .

Additionally, variation with respect to  $\Phi$  gives rise to the generalized PB (GPB) equation:

$$-\nabla \cdot (\epsilon(S) \nabla \Phi) = S \rho_m + (1-S) \sum_{\alpha} q_{\alpha} \rho_{\alpha 0} e^{-\frac{q_{\alpha} \Phi}{k_B T}}, \quad (9)$$

where  $\epsilon(S) = (1-S)\epsilon_s + S\epsilon_m$  is the generalized permittivity function.

## NUMERICAL METHOD

To solve the above coupled equation system, a set of parameters that appeared in the GLB equation, denoted as  $\mathbf{P}$  should be predetermined, the parameter set  $\mathbf{P}$  used in solving the coupled PDEs should meet stability and optimal prediction requirements, our algorithm contains two stages:

- explore the stability conditions of the coupled PDEs by introducing the auxiliary system via a small perturbation;
- optimize the parameter set by an iteratively scheme satisfying the stability constraint.

For the stability concern, we introduced the following constraint:

$$\begin{cases} \frac{\partial S}{\partial t} = |\nabla S| \left[ \nabla \cdot \left( \gamma \frac{\nabla S}{|\nabla S|} \right) - p + U + \frac{1}{2} \epsilon_m |\nabla \Phi|^2 - \frac{1}{2} \epsilon_s |\nabla \Phi|^2 \right], \\ -\nabla \cdot (\epsilon(S) \nabla \Phi) = S \rho_m, \\ \gamma > \gamma_0 > 0, \\ |p| \leq \beta \gamma. \end{cases} \quad (10)$$

Then the parameter optimization problem in the coupled PDEs given by Eqs. (10) can be transformed into the following regularized constrained optimization problem:

$$\min_{\mathbf{P}} \left( \|\Delta G(\mathbf{P}) - \Delta G^{\text{Exp}}\|_2 + \lambda \|\mathbf{P}\|_2 \right), \quad (11)$$

s.t.

$$\gamma \geq \gamma_0, \quad |p| \leq \beta \gamma. \quad (12)$$

A pseudo code given in Algorithm 1 offers a general framework for solving the coupled GLB and GPB equations Eqs.(13-14) in a self-consistent manner.

$$-\nabla \cdot (\epsilon(S) \nabla \Phi) = S \rho_m, \quad (13)$$

and

$$\frac{\partial S}{\partial t} = |\nabla S| \left[ \nabla \cdot \left( \gamma \frac{\nabla S}{|\nabla S|} \right) + V_e \right], \quad (14)$$

where  $V_e$  is the external potential which is defined as:

- **auxiliary system:**  $V_e = \frac{1}{2} (\epsilon_m - \epsilon_s) |\nabla \Phi|^2$ ,
- **full system:**  $V_e = -p + U + \frac{1}{2} (\epsilon_m - \epsilon_s) |\nabla \Phi|^2$ .

**Algorithm 1** Self-consistent algorithm for the coupled GPB and GLB system

```

1: procedure GPB-GLB-SOLVER
2:   Initialize:  $\Delta G_1^{\text{P}} = 0, \Delta G_2^{\text{P}} = 100, \text{Area}_1 = 0, \text{Area}_2 = 100, \text{Vol}_1 = 0, \text{Vol}_2 = 100$ 
3:   do while  $(|\Delta G_1^{\text{P}} - \Delta G_2^{\text{P}}| < \epsilon_1)$ 
4:      $\Delta G_1^{\text{P}} \leftarrow \Delta G_2^{\text{P}}$ 
5:     do while  $(|\text{Area}_1 - \text{Area}_2| < \epsilon_2 \text{ and } |\text{Vol}_1 - \text{Vol}_2| < \epsilon_3)$ 
6:        $\text{Area}_1 \leftarrow \text{Area}_2, \text{Vol}_1 \leftarrow \text{Vol}_2$ .
7:       Update the surface profile function  $S$  by solving the GLB (14).
8:        $\text{Area}_2 = \int_{\Omega} S d\mathbf{r}, \text{Vol}_2 = \int_{\Omega} |\nabla S| d\mathbf{r}$ .
9:     enddo
10:    Solve the GPB (13) in both vacuum and solvent with the previous updated surface profile.
11:    Update the polar solvation free energy  $\Delta G_2^{\text{P}}$ .
12:  enddo
    
```

The parameters  $\epsilon_1, \epsilon_2$ , and  $\epsilon_3$  are the threshold, all set to 0.01 in the implementation.

Algorithm 2 presents the the parameter learning algorithm for a given group of molecules.

**Algorithm 2** Parameters learning for a given group of molecules

```

1: procedure PARAMETERS-LEARNING
2:   Initialize:  $\text{Err}_1 = 0, \text{Err}_2 = 100$ 
3:   Solve the coupled GPB and GLB system, where GLB utilizes the auxiliary equation (14).
4:   Solve the constrained optimization problem Eqs. (11)-(12) to obtain the initial parameters  $\mathbf{P}_0$ .
5:   Update  $\text{Err}_1$  to be the RMS error between experimental and predict results in the above step.
6:   do while  $(|\text{Err}_1 - \text{Err}_2| < \epsilon_4)$ 
7:      $\text{Err}_2 \leftarrow \text{Err}_1$ .
8:     Solve the coupled GPB and GLB, where GLB system with parameters set  $\mathbf{P}_0$ .
9:     Solve the constrained optimization problem Eqs. (11)-(12) to get the updated parameters  $\mathbf{P}$ .
10:    Update  $\text{Err}_1$  to be RMS error between experimental and predict results.
11:    Update  $\mathbf{P}_0 \leftarrow \mathbf{P}$ .
12:  enddo
    
```

The threshold parameter  $\epsilon_4$  is set to 0.01 in the present work.

## NUMERICAL RESULTS

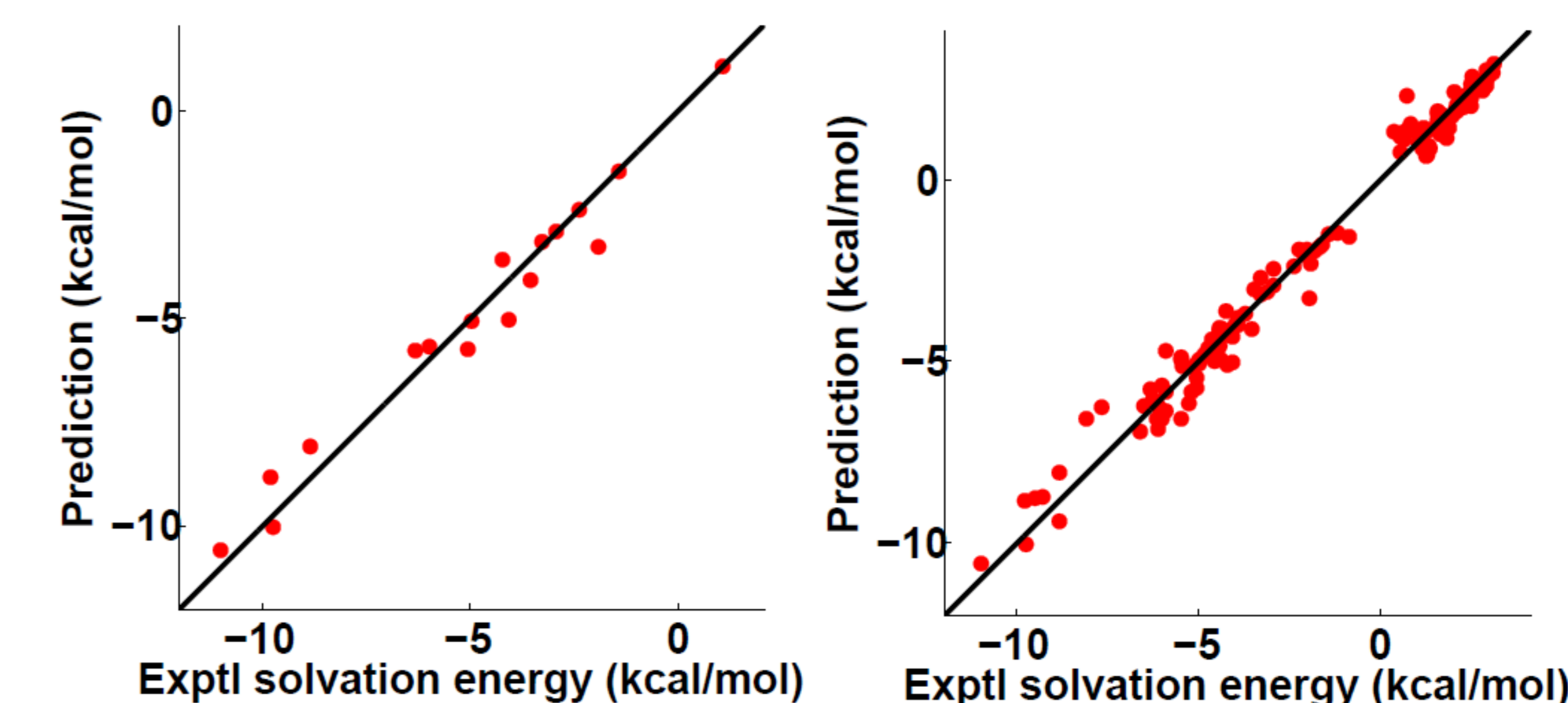


Figure 1. The comparison of predicted and experimental solvation free energy for SAMPL0 set (Left chart) and 138 molecules (Right chart). Their RMS errors are 0.60 kcal/mol and 0.48 kcal/mol, respectively.

Figures 2 and 3 demonstrate the five-fold cross validations for four types of molecules.

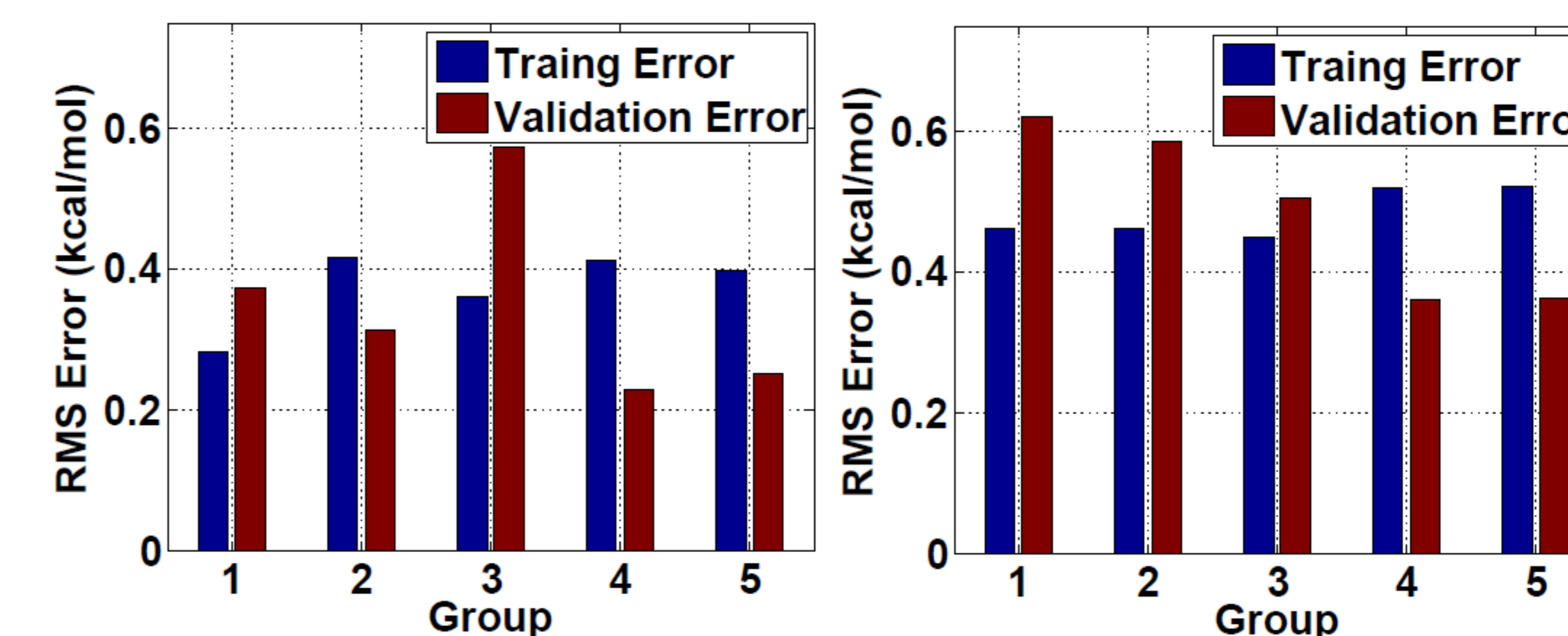


Figure 2. Five-fold cross validations for 38 alkane (Left chart) and 23 alkene (Right chart) molecules.

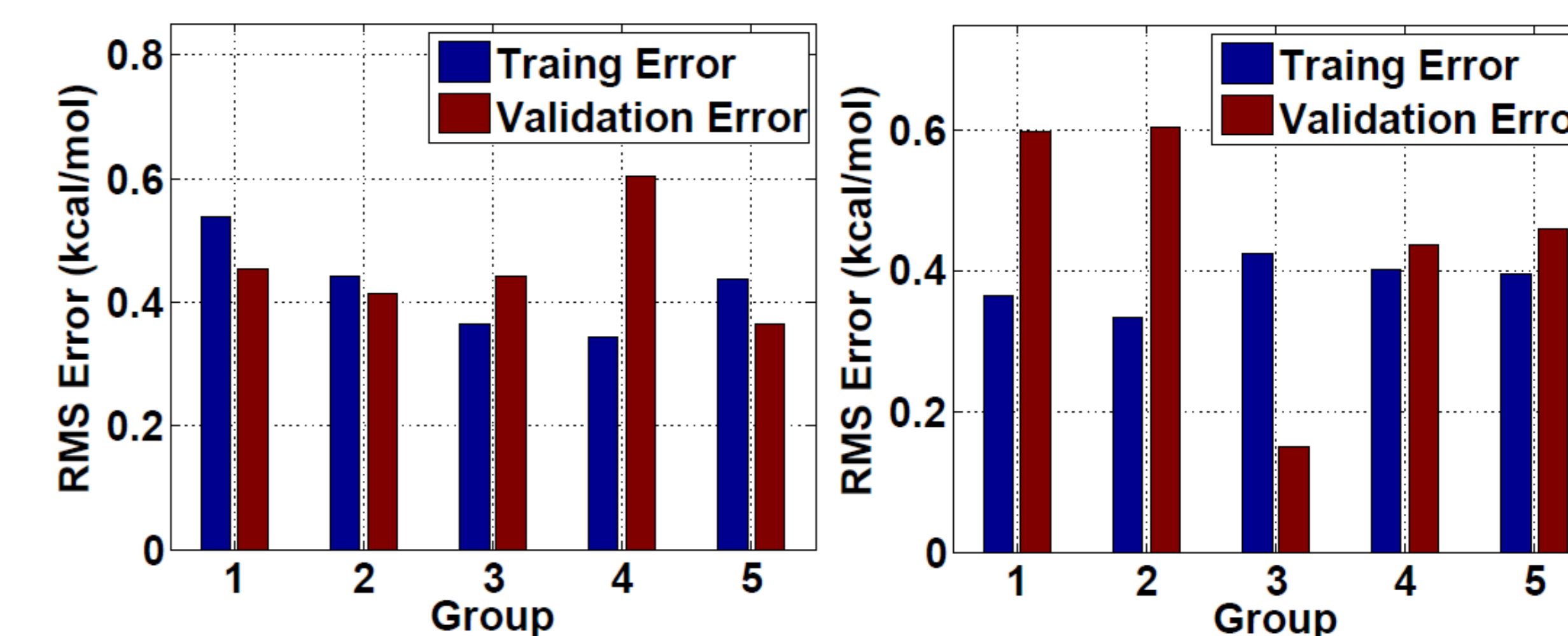


Figure 3. Five-fold cross validations for 25 alcohol (Left chart) and 18 ether (Right chart) molecules.

## REFERENCES

1. Bao Wang, Nathan A. Baker, Guo-Wei Wei, Parameter Optimization in the Differential Geometry based Solvation Models, preprint, 2015.
2. Zhan Chen, Nathan A. Baker, G.W. Wei, Differential geometry based solvation model I: Eulerian formulation, Journal of Computational Physics 229 (2010) 8231C8258.

## 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.

