

# Toward a Universal Model To Calculate the Solvation Thermodynamics of Druglike Molecules: The Importance of New Experimental Databases

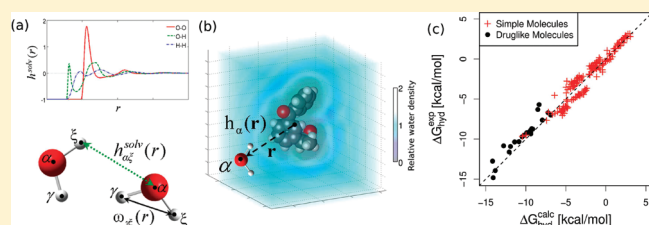
David S. Palmer, Andrey I. Frolov, Ekaterina L. Ratkova, and Maxim V. Fedorov\*

Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, DE-04103 Leipzig, Germany

**S** Supporting Information

**ABSTRACT:** We demonstrate that a new free energy functional in the integral equation theory of molecular liquids gives accurate calculations of hydration thermodynamics for druglike molecules. The functional provides an improved description of excluded volume effects by incorporating two free coefficients. When the values of these coefficients are obtained from experimental data for simple organic molecules, the hydration free energies of an external test set of druglike molecules can be calculated with an accuracy of about 1 kcal/mol. The 3D RISM/UC method proposed here is easily implemented using existing computational software and allows in silico screening of the solvation thermodynamics of potential pharmaceutical molecules at significantly lower computational expense than explicit solvent simulations.

**KEYWORDS:** hydration free energy, reference interaction site model, RISM, solubility, drug discovery, RISM/UC, bioavailability, in silico, virtual, screening



## INTRODUCTION

The integral equation theory (IET) of molecular liquids is a theoretical framework for modeling solute–solvent effects based upon the molecular Ornstein–Zernike (MOZ) equation. IET has previously been used in *qualitative* studies of solution chemistry phenomena ranging from solvation of monatomic ions to solvent effects on biomolecules and supramolecular assemblies.<sup>1–8</sup> The theory allows the calculation of solvation thermodynamics and solvent structure at significantly lower computational expense than molecular dynamics simulations. Although IET has been an active topic of academic research for over 40 years, in its current state it does not permit accurate calculations of solvation thermodynamics across multiple classes of molecules, which has prevented it from being widely used in many practical applications such as pharmaceutical research and development.<sup>9,10</sup> The purpose of this article is to show that the hydration free energies (HFE) of molecules ranging from simple alkanes to pharmaceuticals can be calculated accurately using a new free energy functional in the scope of the 3D reference interaction site model (an IET method). Although the new functional is in an early stage of development, the results are already as accurate as those obtained from well-developed implicit solvent models that contain a large number of empirical parameters. Unlike the continuum solvent models, however, the IET model retains information about the solvent structure and can be used to probe specific solute–solvent interactions.

The hydration free energy (HFE) is a key thermodynamic parameter in characterizing a solute–solvent system. Many physicochemical properties of molecules are defined by their solvation and acid–base behavior, which can be estimated from their HFEs. For example, HFEs have been used in the calculation of acid–base

dissociation constants ( $pK_a$ 's),<sup>11</sup> aqueous solubilities,<sup>12,13</sup> octanol–water partition coefficients,<sup>14–16</sup> and protein–ligand binding free energies.<sup>17,18</sup> As these physicochemical properties are used in predicting the pharmacokinetic behavior of novel pharmaceutical molecules,<sup>10</sup> improving the accuracy of computational methods to calculate HFEs would have widespread benefits.

## THEORY

**Background.** Integral equation theory (IET) is based on the molecular Ornstein–Zernike (MOZ) integral equation of classical density functional theory, which operates with binary spatial density correlation functions. Due to the high-dimensionality of the MOZ equations, however, a number of approximations must be introduced to allow them to be solved for systems of chemical interest.<sup>2,20</sup> Here we work with the 3D reference interaction site model (3D RISM),<sup>2,4,21,22</sup> which has previously been demonstrated to be an effective method for qualitative modeling of many different solution phase phenomena.<sup>2,4,22</sup> The 3D RISM equations relate 3D intermolecular *solvent site–solute* total correlation functions ( $h_\alpha(\mathbf{r})$ ) and direct correlation functions ( $c_\alpha(\mathbf{r})$ ) (index  $\alpha$  corresponds to the solvent sites) (Figure 1):<sup>2,21</sup>

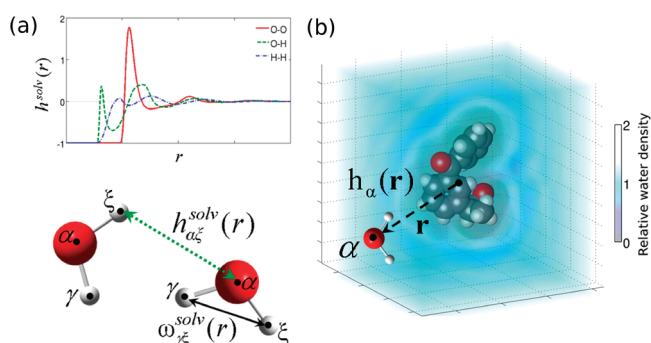
$$h_\alpha(\mathbf{r}) = \sum_{\xi=1}^{N_{\text{solvent}}} \int_{R^3} c_\xi(\mathbf{r}-\mathbf{r}') \chi_{\xi\alpha}(|\mathbf{r}'|) d\mathbf{r}' \quad (1)$$

**Received:** March 11, 2011

**Accepted:** May 27, 2011

**Revised:** May 6, 2011

**Published:** May 27, 2011



**Figure 1.** Correlation functions in the 3D RISM approach. (a) Site–site intramolecular ( $\omega_{\alpha\beta}^{\text{sol}}(r)$ ) and intermolecular ( $h_{\alpha\beta}^{\text{sol}}(r)$ ) correlation functions between sites of solvent molecules. The inset shows the radial projections of solvent site–site density correlation functions: oxygen–oxygen (OO, red solid), oxygen–hydrogen (OH, green dashed) and hydrogen–hydrogen (HH, blue dash-dotted). (b) Three-dimensional intermolecular solute–solvent correlation function  $h_{\alpha}(\mathbf{r})$  around a model solute. This figure is based on and in part reproduces Figure 1 from our earlier work<sup>19</sup> and is reprinted with permission. Copyright 2010 IOP Publishing Ltd.<sup>19</sup>

where  $\chi_{\xi\alpha}(r)$  is the bulk solvent susceptibility function, and  $N_{\text{solvent}}$  is the number of sites in a solvent molecule. The solvent susceptibility function  $\chi_{\xi\alpha}(r)$  describes the mutual correlations of sites  $\xi$  and  $\alpha$  in solvent molecules in the bulk solvent. It can be obtained from the solvent intramolecular correlation function ( $\omega_{\xi\alpha}^{\text{sol}}(r)$ ), site–site radial total correlation functions ( $h_{\xi\alpha}^{\text{sol}}(r)$ ) and the solvent site number density ( $\rho_{\alpha}$ ):  $\chi_{\xi\alpha}(r) = \omega_{\xi\alpha}^{\text{sol}}(r) + \rho h_{\xi\alpha}^{\text{sol}}(r)$  (Figure 1a).<sup>2</sup> In this work, these functions were obtained by solution of the 1D RISM equations of the pure solvent.<sup>2,23</sup>

To make eq 1 complete,  $N_{\text{solvent}}$  closure relations are introduced:

$$h_{\alpha}(\mathbf{r}) = \exp(-\beta u_{\alpha}(\mathbf{r}) + h_{\alpha}(\mathbf{r}) - c_{\alpha}(\mathbf{r}) + B_{\alpha}(\mathbf{r})) - 1$$

$$\alpha = 1, \dots, N_{\text{solvent}} \quad (2)$$

where  $u_{\alpha}(\mathbf{r})$  is the 3D interaction potential between the solute molecule and  $\alpha$  solvent site,  $B_{\alpha}(\mathbf{r})$  are bridge functionals,  $\beta = 1/k_{\text{B}}T$ ,  $k_{\text{B}}$  is the Boltzmann constant, and  $T$  is the temperature.

In general, the exact bridge functions  $B_{\alpha}(\mathbf{r})$  in eq 2 are represented as an infinite series of integrals over high order correlation functions and are therefore practically incomputable, which makes it necessary to incorporate some approximations.<sup>1,2,24</sup> In the current work, we use a closure relationship proposed by Kovalenko and Hirata (the KH closure),<sup>25</sup> which was designed to improve convergence rates and to prevent possible divergence of the numerical solution of the RISM equations:<sup>25</sup>

$$h_{\alpha}(\mathbf{r}) = \begin{cases} \exp(\Xi_{\alpha}(\mathbf{r})) - 1 & \text{when } \Xi_{\alpha}(\mathbf{r}) < 0 \\ \Xi_{\alpha}(\mathbf{r}) & \text{when } \Xi_{\alpha}(\mathbf{r}) > 0 \end{cases} \quad (3)$$

where  $\Xi_{\alpha}(\mathbf{r}) = -\beta u_{\alpha}(\mathbf{r}) + h_{\alpha}(\mathbf{r}) - c_{\alpha}(\mathbf{r})$ .

The 3D interaction potential between the solute molecule and  $\alpha$  site of solvent ( $u_{\alpha}(\mathbf{r})$ , eq 2) is estimated as a superposition of the site–site interaction potentials between solute sites and the particular solvent site, which depend only on the absolute distance between the two sites. We use the common form of the site–site interaction potential represented by the long-range electrostatic interaction term and the short-range term (Lennard-Jones potential).<sup>26</sup> The Lennard-Jones parameters on atom sites were taken from

the general amber force field (GAFF),<sup>27</sup> and partial charges on atoms were computed with the AM1-BCC approach<sup>27–30</sup> as was done in our previous work.<sup>19</sup>

**New Free Energy Functional.** Within the framework of the RISM theory there exist several approximate functionals that allow one to analytically obtain values of the HFE from the total  $h_{\alpha}(\mathbf{r})$  and direct  $c_{\alpha}(\mathbf{r})$  correlation functions.<sup>18,31,32</sup> Although these functionals have been extensively used to *qualitatively* model thermodynamics of different chemical systems,<sup>18,33,34</sup> they generally give HFE values that are strongly biased from experimental data with a large standard deviation error.<sup>18,26,31,32,35,36</sup> In this work, we show that the hydration free energies (HFE) of molecules ranging from simple alkanes to pharmaceuticals can be calculated accurately using a new free energy functional in the scope of the 3D reference interaction site model.

We consider a HFE functional initially developed by Chandler, Singh and Richardson, and adopted by Kovalenko and Hirata for the 3D RISM case, which assumes Gaussian fluctuations (GF) of the solvent:<sup>2,18,37</sup>

$$\Delta G_{\text{hyd}}^{\text{GF}} = k_{\text{B}}T \sum_{\alpha=1}^{N_{\text{solvent}}} \rho_{\alpha} \int_{R^3} \left[ -c_{\alpha}(\mathbf{r}) - \frac{1}{2} c_{\alpha}(\mathbf{r}) h_{\alpha}(\mathbf{r}) \right] d\mathbf{r} \quad (4)$$

where  $\rho_{\alpha}$  is the number density of a solvent sites  $\alpha$ . We use this HFE functional here because in our recent works<sup>26,31</sup> we found that it performs better for HFE calculations than traditionally used HNC<sup>38</sup> or KH<sup>25</sup> functionals for RISM HFEs. Although RISM works with a realistic model of solvent structure, it has previously been suggested that there is a systematic error in the cavity formation term calculated (implicitly) by the theory.<sup>35,39,40</sup> For example, the error in RISM calculations of HFEs has previously been attributed to an overestimation of the energy required to form a cavity in the solvent.<sup>35,39,41</sup> Kinoshita and co-workers have also highlighted the importance of the excluded volume contribution to the hydration free energy in the morphometric approach to the solvation free energy<sup>40</sup> and works on the hydration thermodynamics of biomolecules.<sup>42</sup> We have recently shown that the partial molar volume calculated by RISM is well correlated with the difference between experimental and RISM calculated hydration free energies for simple organic molecules,<sup>19,31</sup> which suggests that it can be used with a linear scaling coefficient as a universal correction (UC) for the 3D RISM calculations. Therefore, we define a free energy functional which is a linear combination of the  $\Delta G_{\text{hyd}}^{\text{GF}}$ , the dimensionless partial molar contribution,  $\rho V$ , and a bias correction,  $b$  (intercept):<sup>19</sup>

$$\Delta G_{\text{hyd}}^{\text{3D-RISM/UC}} = \Delta G_{\text{hyd}}^{\text{GF}} + a(\rho V) + b \quad (5)$$

where the scaling coefficient  $a$  and intercept  $b$  values are obtained by linear regression against the experimental data for the simple organic molecule data set. For the combination of methods used here (e.g., KH closure, GF free energy functional, molecular geometries optimized at the AM1 level of theory, AM1-BCC partial charges, and Lennard-Jones parameters taken from the AMBER GAFF forcefield), the coefficients have the values  $a = -3.2217$  kcal/mol and  $b = 0.5783$  kcal/mol (more details are provided below).

We estimate the solute partial molar volume via *solute–solvent* site correlation functions using the standard 3D RISM theory expression:<sup>43,44</sup>

$$V = k_{\text{B}}T\eta \left( 1 - \rho_{\alpha} \sum_{\alpha=1}^{N_{\text{solvent}}} \int_{R^3} c_{\alpha}(\mathbf{r}) d\mathbf{r} \right) \quad (6)$$

where  $\eta$  is the pure solvent isothermal compressibility and  $\rho_\alpha$  is the number density of solute sites  $\alpha$ .

## METHODS

**Data Sets.** To demonstrate the efficiency of the method, and the transferability of the parameters, we have used it to calculate HFEs for two data sets of molecules: (i) 163 neutral simple organic molecules taken from our previous work<sup>31</sup> and (ii) 21 druglike molecules in their neutral form with experimental HFE data extracted from the literature. The majority of the data in our druglike test set were taken from publications by Perlovich and Bauer-Brandl.<sup>45–50</sup> We note that the lack of accurate and well-documented experimental hydration free energy data for druglike molecules in the published literature is a significant stumbling block in the development of new computational solvent models.<sup>51,52</sup> (By “well-documented” we mean that both the methodology and the experimental conditions must be clearly reported.) The experimental HFE data are given here as  $\Delta G_{\text{hyd}} = -RT \ln c_{\text{aq}}/c_{\text{gas}}$ , with HFEs in kcal/mol and concentrations in mol/L, which corresponds to the choice of standard states suggested by Ben-Naim.<sup>53,54</sup> The experimental and calculated hydration free energy data for the druglike molecules are provided in the Supporting Information. We have chosen not to use the data sets published as part of the SAMPL1<sup>55</sup> and SAMPL2<sup>51</sup> challenges because these contain a large number of pesticides and we are interested in assessing our methods on druglike molecules.

**Computational Details. RISM Calculations.** RISM calculations were performed assuming a temperature of 300 K and infinitely diluted solution. We used the Lue and Blankshtein version of the SPC/E model of water (MSPC/E).<sup>56</sup> This differs from the original SPC/E water model<sup>57</sup> by the addition of modified LJ potential parameters for the water hydrogen, which were altered to prevent possible divergence of the algorithm.<sup>41,58–60</sup> The Lorentz–Berthelot mixing rules were used to generate the solute–water LJ potential parameters.<sup>61</sup> The following LJ parameters (for water hydrogen) were used to calculate the interactions between solute sites and water hydrogens:  $\sigma_{\text{H}_w}^{\text{LJ}} = 1.1657 \text{ \AA}$  and  $\epsilon_{\text{H}_w}^{\text{LJ}} = 0.0155 \text{ kcal/mol}$ .

**3D RISM Calculations.** The 3D RISM calculations were performed using the NAB simulation package<sup>23,62–64</sup> in the AmberTools 1.4 set of routines.<sup>65</sup> The 3D-grid around a solute was generated such that the minimum distance between any solute atom and the edge of solvent box (*buffer* in NAB notation) was equal to 30 Å. The linear grid spacing in each of the three directions was 0.3 Å. We employed the MDIIS iterative scheme,<sup>66</sup> where we used 5 MDIIS vectors, MDIIS step size  $-0.7$ , residual tolerance  $-10^{-10}$ . The KH closure was used for solution of the 3D RISM equations. Solvent susceptibility functions for pure water were taken from the 1D RISM calculations.

**Solvent Susceptibility Functions.** Solvent susceptibility functions were calculated with the 1D RISM method present in AmberTools 1.4. The dielectrically consistent RISM was employed,<sup>67</sup> using the KH closure. The grid size for 1D-functions was 0.025 Å, which gave a total of 16384 grid points. We employed the MDIIS iterative scheme, where we used 20 MDIIS vectors, MDIIS step size  $-0.3$ , and residual tolerance  $-10^{-12}$ . The solvent was considered to be pure water with a number density  $0.0333 \text{ \AA}^{-3}$  and a dielectric constant of 78.497, at a temperature of 300 K. The final susceptibility solvent site–site functions were stored and then used as input for the 3D RISM

calculations. The solvent isothermal compressibility evaluated from the 1D RISM calculation was  $k_B T \eta = 1.949459 \text{ \AA}^3$ .

**Input Structures and Potential Parameters.** The following data are needed for 3D RISM calculations in the NAB simulation package: (1) atomic coordinates, (2) partial charges of atoms, and (3) atom–atom potential parameters representing the van der Waals interactions. These parameters were assigned to each molecule using programs distributed with the AmberTools 1.4 package.<sup>65,68,69</sup>

- (1) The coordinates of each molecule were optimized using the AM1 Hamiltonian<sup>30</sup> via the *antechamber*<sup>70</sup> suite, which uses the *sqm*<sup>65</sup> program for semiempirical QM calculations. For the data set of simple molecules, the initial configurations for these QM geometry optimizations were taken from our previous work.<sup>31</sup> For the data set of druglike molecules, a low-mode conformational search<sup>71,72</sup> was carried out in vacuum using the OPLS2005 forcefield in MacroModel v.9.1.<sup>73</sup> The global minimum energy conformer for each molecule was then used as input for the AM1 structure optimizations.
- (2) Atomic partial charges were calculated using the AM1-BCC method,<sup>27–29</sup> where BCC stands for bond charge correction, as implemented in Antechamber from the AmberTools 1.4 package.<sup>65</sup> The BCC parameters were taken from Jakalian et al.<sup>29</sup>
- (3) For all compounds, the LJ parameters from the general amber force field (GAFF)<sup>27</sup> were assigned to solute atoms with the *antechamber* and *tleap* programs.<sup>70</sup>

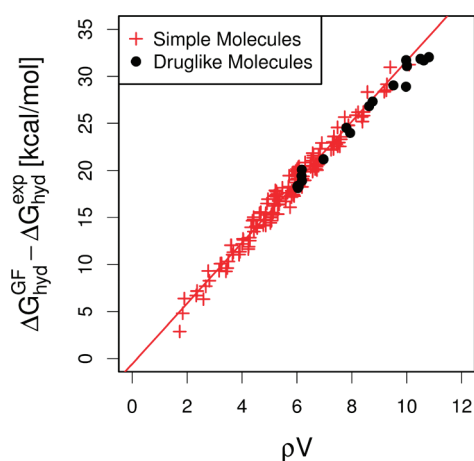
**Computational Expense.** The calculations discussed in this paper were performed on an Intel Core 2 Duo CPU E8600 3.33 GHz processor. The mean time required to calculate the hydration free energy of a single solute from the training set was  $\sim 30$  min, while the minimum and maximum values were  $\sim 12$  min and  $\sim 45$  min, respectively, depending on the size of the molecule in question. For the druglike molecules in the test set, the calculations took between  $\sim 30$  min and  $\sim 75$  min with a mean value of  $\sim 45$  min. We note that the time required for a single calculation might be significantly reduced by using advanced numerical algorithms<sup>74</sup> or by performing the simulations using parallel computation.<sup>23</sup>

## RESULTS AND DISCUSSION

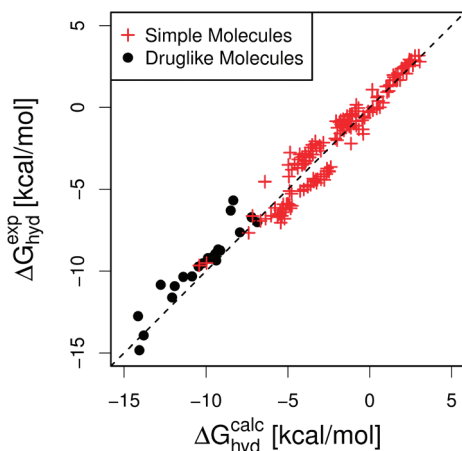
The HFEs calculated by 3D RISM using the GF free energy functional have only a weak correlation with the experimental data for both the simple molecule data set ( $R = 0.55$ , RMSE = 18.51 kcal/mol,  $\sigma = 5.14 \text{ kcal/mol}$ , bias = 17.78 kcal/mol) and the druglike data set ( $R = 0.68$ , RMSE = 25.41 kcal/mol,  $\sigma = 5.50 \text{ kcal/mol}$ , bias = 24.81 kcal/mol. Correlation plots for both data sets are given in the Supporting Information). The bias in the calculated HFEs indicates that 3D RISM predicts these molecules to be too hydrophobic. However, when the error in the calculated HFE is plotted against the calculated partial molar volume of the solute, a very high linear correlation is observed for both the simple molecule data set (Figure 2,  $R = 0.99$ ) and the druglike data set (Figure 2,  $R = 0.99$ ). Moreover, the correlation is equally high for all classes of molecules considered here. For instance, in Figure 2, it is shown that the druglike molecules lie on the line-of-best-fit calculated for the simple organic molecules.

Since the correlation observed in Figure 2 is high for all classes of molecules, the model to predict the HFEs of drug molecules (eq 5) can be trained on the simple, organic molecules, thereby





**Figure 2.** Correlation between the error in the hydration free energies calculated by 3D-RISM-KH theory using the Gaussian fluctuations free energy functional ( $\Delta G_{\text{hyd}}^{\text{GF}} - \Delta G_{\text{hyd}}^{\text{exp}}$ ) and the calculated partial molar volume ( $\rho V$ ). Simple organic molecules are given as red crosses. Druglike molecules are given as black circles.



**Figure 3.** Correlation between experimental ( $\Delta G_{\text{hyd}}^{\text{exp}}$ ) and calculated ( $\Delta G_{\text{hyd}}^{\text{calc}}$ ) hydration free energies, where the calculated values are obtained using the 3D-RISM/UC model proposed here. Simple organic molecules are given as red crosses. Druglike molecules are given as black circles.

demonstrating the transferability of the two free model coefficients (we refer to the model as 3D-RISM/UC). The values of the empirical coefficients  $a$  and  $b$  obtained by linear regression against the data set of simple, organic molecules are  $-3.2217$  and  $0.5783$  kcal/mol, respectively, where the regression statistics are  $R = 0.96$ ,  $\text{RMSE} = 0.82$  kcal/mol,  $\sigma = 0.82$  kcal/mol,  $\text{bias} = 0.00$  kcal/mol. The model gives very accurate calculations of the HFEs of the data set of pharmaceuticals (Figure 3), where the statistics are  $R = 0.94$ ,  $\text{RMSE} = 1.06$  kcal/mol,  $\sigma = 0.78$  kcal/mol,  $\text{bias} = -0.72$  kcal/mol.

To benchmark these results, we have calculated HFEs for the 21 druglike molecules using three commonly used continuum solvent models in the scope of quantum mechanics: (1, 2) the SM8<sup>75,76</sup> and SM6<sup>77</sup> models using the MO6-2X density functional and 6-31G(d) basis set as implemented in Jaguar;<sup>78</sup> (3) the polarizable continuum model using HF theory and the 6-31G(d) basis set combined with United Atom for Hartree-Fock (UAHF) atomic radii as implemented in Gaussian (further details of these

**Table 1.** Statistics for Prediction of the Hydration Free Energies of 21 Druglike Molecules Using 3D-RISM/UC, HF/6-31G(d) PCM, MO6-2X/6-31G(d) SM8, and MO6-2X/6-31G(d) SM6

method	R	RMSE (kcal/mol)	$\sigma$ (kcal/mol)	bias (kcal/mol)
3DRISM/UC	0.94	1.06	0.78	-0.72
HF/6-31G(d) PCM	0.96	1.12	0.65	0.91
MO6-2X/6-31G(d) SM8	0.94	0.82	0.79	-0.21
MO6-2X/6-31G(d) SM6	0.94	1.16	0.84	0.8

calculations are provided in the Supporting Information).<sup>79</sup> These combinations of QM theory and solvent model were selected because they performed well in a recent blind challenge for HFE calculation.<sup>80–82</sup> We note that these are also the recommended methods for HFE calculation in Jaguar and Gaussian, respectively. The results of these calculations are given in Table 1. It is clear that even without additional parametrization the 3D-RISM/UC model gives excellent estimates of HFEs that are already as accurate as those obtained from leading implicit solvent models. Unlike the continuum solvent models, however, the IET model can in principle be used to probe specific solute–solvent effects since it retains information about the solvent structure in terms of total correlation functions ( $h_{\alpha}(\mathbf{r})$ ). Correlation functions calculated by 3D-RISM have previously been used in fragment-based drug design,<sup>4</sup> modeling the binding of water<sup>6</sup> and ions<sup>5</sup> by proteins, and interpreting solvent densities around biomacromolecules.<sup>83,84</sup>

## CONCLUSIONS

We have shown that accurate calculations of hydration free energies can be obtained using the 3D-RISM/UC approach, which is based on 3D-RISM-KH theory combined with the GF free energy functional and a correction based on the partial molar volume of the solute. Moreover, the empirical coefficients in the 3D-RISM/UC model can be calculated for simple organic molecules and used to give accurate predictions for druglike molecules, which demonstrates that they are transferable between chemical classes. For an external test set containing 21 druglike molecules from different chemical classes, the hydration free energies calculated using the 3D-RISM/UC model were found to be in good agreement with experimental data, with a standard deviation of the error of approximately 1 kcal/mol, which even without additional parametrization is comparable to the results of long-standing and well developed implicit continuum solvent models. The results show that the 3D-RISM/UC free energy functional gives accurate calculations of HFEs across multiple classes of molecules ranging from simple organic molecules to different classes of druglike molecules. Whether this trend remains the same for the whole of druglike chemical space cannot at present be assessed because of a lack of experimental HFE data for druglike molecules. Moreover, further work is required to assess the accuracy of the 3D-RISM/UC approach for biological macromolecules and to determine whether it can be used to calculate hydration terms in the prediction of protein–ligand binding free energies.<sup>18</sup> Nevertheless, we believe that the model proposed here opens up new avenues for research in the computational physical chemistry of solutions. The current limiting step in development of the model is a lack of accurate experimental thermodynamic data for pharmaceutical and other important classes of molecules. We hope that our findings demonstrate the need for new computational databases of experimental

solvation thermodynamic data for molecules taken from the full extent of druglike chemical space.

## ■ ASSOCIATED CONTENT

**S Supporting Information.** The complete data sets including all experimental and calculated data and details of conversions of standard state. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, DE-04103 Leipzig, Germany. E-mail: [fedorov@mis.mpg.de](mailto:fedorov@mis.mpg.de). Tel: +49 341 9959 756. Fax: +49 341 9959 804.

## ■ ACKNOWLEDGMENT

We acknowledge the partial financial support of the Deutsche Forschungsgemeinschaft (DFG), German Research Foundation, Research Grant FE 1156/2-1, and also of Grant No 247500-(BioSol) of the European Commission, REA, Research Executive Agency, Programme: FP7-PEOPLE-2009-IRSES. We are grateful for use of the computing facilities at the John von Neumann-Institut für Computing (NIC), Juelich Supercomputing Centre (JSC), Forschungszentrum Juelich GmbH, Germany; Project IDs: HLZ18 and HLZ16.

## ■ REFERENCES

- (1) Hansen, J.-P.; McDonald, I. R. *Theory of Simple Liquids*, 4th ed; Elsevier Academic Press: Amsterdam, The Netherlands, 2000.
- (2) Hirata, F., Ed. *Molecular theory of solvation*; Kluwer Academic Publishers: Dordrecht, Netherlands, 2003.
- (3) Kiyota, Y.; Hiraoka, R.; Yoshida, N.; Maruyama, Y.; Imai, I.; Hirata, F. Theoretical Study of CO Escaping Pathway in Myoglobin with the 3D-RISM Theory. *J. Am. Chem. Soc.* **2009**, *131*, 3852–3853.
- (4) Imai, T.; Oda, K.; Kovalenko, A.; Hirata, F.; Kidera, A. Ligand Mapping on Protein Surfaces by the 3D-RISM Theory: Toward Computational Fragment-Based Drug Design. *J. Am. Chem. Soc.* **2009**, *131*, 12430–12440.
- (5) Yoshida, N.; Phongphanphanee, S.; Maruyama, Y.; Imai, T.; Hirata, F. Selective ion-binding by protein probed with the 3D-RISM theory. *J. Am. Chem. Soc.* **2006**, *128*, 12042–12043.
- (6) Imai, T.; Hiraoka, R.; Kovalenko, A.; Hirata, F. Water molecules in a protein cavity detected by a statistical-mechanical theory. *J. Am. Chem. Soc.* **2005**, *127*, 15334–15335.
- (7) Kinoshita, M.; Okamoto, Y.; Hirata, F. Peptide conformations in alcohol and water: Analyses by the reference interaction site model theory. *J. Am. Chem. Soc.* **2000**, *122*, 2773–2779.
- (8) Kinoshita, M.; Okamoto, Y.; Hirata, F. First-principle determination of peptide conformations in solvents: Combination of Monte Carlo simulated annealing and RISM theory. *J. Am. Chem. Soc.* **1998**, *120*, 1855–1863.
- (9) Orozco, M.; Luque, F. J. Theoretical methods for the description of the solvent effect in biomolecular systems. *Chem. Rev.* **2000**, *100*, 4187–4225.
- (10) Jorgensen, W. L. The many roles of computation in drug discovery. *Science* **2004**, *303*, 1813–1818.
- (11) Jensen, J. H.; Li, H.; Robertson, A. D.; Molina, P. A. Prediction and rationalization of protein pK(a) values using QM and QM/MM methods. *J. Phys. Chem. A* **2005**, *109*, 6634–6643.
- (12) Jorgensen, W. L.; Duffy, E. M. Prediction of drug solubility from structure. *Adv. Drug Delivery Rev.* **2002**, *54*, 355–366.
- (13) Palmer, D. S.; Llinas, A.; Morao, I.; Day, G. M.; Goodman, J. M.; Glen, R. C.; Mitchell, J. B. O. Predicting intrinsic aqueous solubility by a thermodynamic cycle. *Mol. Pharmaceutics* **2008**, *5*, 266–279.
- (14) Garrido, N. M.; Queimada, A. J.; Jorge, M.; Macedo, E. A.; Economou, I. G. 1-Octanol/Water Partition Coefficients of n-Alkanes from Molecular Simulations of Absolute Solvation Free Energies. *J. Chem. Theory Comput.* **2009**, *5*, 2436–2446.
- (15) Hughes, L. D.; Palmer, D. S.; Nigsch, F.; Mitchell, J. B. O. Why are some properties more difficult to predict than others? A study of QSPR models of solubility, melting point, and Log P. *J. Chem. Inf. Model.* **2008**, *48*, 220–232.
- (16) Perlovich, G.; Bauer-Brandl, A. Solvation of Drugs as a Key for Understanding Partitioning and Passive Transport Exemplified by NSAIDs. *Curr. Drug Delivery* **2004**, *1*, 213–226.
- (17) Swanson, J. M. J.; Henchman, R. H.; McCammon, J. A. Revisiting free energy calculations: A theoretical connection to MM/PBSA and direct calculation of the association free energy. *Biophys. J.* **2004**, *86*, 67–74.
- (18) Genheden, S.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Ryde, U. An MM/3D-RISM Approach for Ligand Binding Affinities. *J. Phys. Chem. B* **2010**, *114*, 8505–8516.
- (19) Palmer, S.; Frolov, A. I.; Ratkova, E. L.; Fedorov, M. V. Towards a universal method to calculate hydration free energies: a 3D reference interaction site model with partial molar volume correction. *J. Phys.: Condens. Matter* **2010**, *22*, 492101.
- (20) Chandler, D.; Andersen, H. C. Optimized cluster expansions for classical fluids. 2. Theory of molecular liquids. *J. Chem. Phys.* **1972**, *57*, 1930–1937.
- (21) Beglov, D.; Roux, B. An integral equation to describe the solvation of polar molecules in liquid water. *J. Phys. Chem.* **1997**, *101*, 7821–7826.
- (22) Du, Q. H.; Beglov, D.; Roux, B. Solvation free energy of polar and nonpolar molecules in water: An extended interaction site integral equation theory in three dimensions. *J. Phys. Chem. B* **2000**, *104*, 796–805.
- (23) Luchko, T.; Gusarov, S.; Roe, D. R.; Simmerling, C.; Case, D. A.; Tuszynski, J.; Kovalenko, A. Three-Dimensional Molecular Theory of Solvation Coupled with Molecular Dynamics in Amber. *J. Chem. Theory Comput.* **2010**, *6*, 607–624.
- (24) Duh, D. M.; Haymet, A. D. J. Integral-Equation Theory For Uncharged Liquids: The Lennard-Jones Fluid And The Bridge Function. *J. Chem. Phys.* **1995**, *103*, 2625–2633.
- (25) Kovalenko, A.; Hirata, F. Potential of mean force between two molecular ions in a polar molecular solvent: A study by the three-dimensional reference interaction site model. *J. Phys. Chem. B* **1999**, *103*, 7942–7957.
- (26) Palmer, D. S.; Sergiievskiy, V. P.; Jensen, F.; Fedorov, M. V. Accurate calculations of the hydration free energies of druglike molecules using the reference interaction site model. *J. Chem. Phys.* **2010**, *133*, 044104.
- (27) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (28) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic Charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132–146.
- (29) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *23*, 1623–1641.
- (30) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. The Development And Use Of Quantum-Mechanical Molecular-Models 0.76. Am1 - A New General-Purpose Quantum-Mechanical Molecular-Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
- (31) Ratkova, E. L.; Chuev, G. N.; Sergiievskiy, V. P.; Fedorov, M. V. An Accurate Prediction of Hydration Free Energies by Combination of Molecular Integral Equations Theory with Structural Descriptors. *J. Phys. Chem. B* **2010**, *114*, 12068–12079.
- (32) Ten-no, S.; Jung, J.; Chuman, H.; Kawashima, Y. Assessment of free energy expressions in RISM integral equation theory: theoretical

prediction of partition coefficients revisited. *Mol. Phys.* **2010**, *108*, 327–332.

(33) Drabik, P.; Gusarov, S.; Kovalenko, A. Microtubule stability studied by three-dimensional molecular theory of solvation. *Biophys. J.* **2007**, *92*, 394–403.

(34) Blinov, N.; Dorosh, L.; Wishart, D.; Kovalenko, A. Association Thermodynamics and Conformational Stability of beta-Sheet Amyloid beta(17–42) Oligomers: Effects of E22Q (Dutch) Mutation and Charge Neutralization. *Biophys. J.* **2010**, *98*, 282–296.

(35) Ten-no, S. Free energy of solvation for the reference interaction site model: Critical comparison of expressions. *J. Chem. Phys.* **2001**, *115*, 3724–3731.

(36) Ratkova, E. L.; Fedorov, M. V. Combination of RISM and Cheminformatics for Efficient Predictions of Hydration Free Energy of Polyfragment Molecules: Application to a Set of Organic Pollutants. *J. Chem. Theory Comput.* **2011**, *7*, 1450–1457.

(37) Chandler, D.; Singh, Y.; Richardson, D. M. Excess Electrons In Simple Fluids 0.1. General Equilibrium-Theory For Classical Hard-Sphere Solvents. *J. Chem. Phys.* **1984**, *81*, 1975–1982.

(38) Singer, S. J.; Chandler, D. Free-energy Functions in the Extended Rism Approximation. *Mol. Phys.* **1985**, *55*, 621–625.

(39) Sato, K.; Chuman, H.; Ten-no, S. Comparative study on solvation free energy expressions in reference interaction site model integral equation theory. *J. Phys. Chem. B* **2005**, *109*, 17290–17295.

(40) Roth, R.; Harano, Y.; Kinoshita, M. Morphometric Approach to the Solvation Free Energy of Complex Molecules. *Phys. Rev. Lett.* **2006**, *97*, 078101.

(41) Chuev, G.; Fedorov, M.; Crain, J. Improved estimates for hydration free energy obtained by the reference interaction site model. *Chem. Phys. Lett.* **2007**, *448*, 198–202.

(42) Kinoshita, M. Roles of translational motion of water molecules in sustaining life. *Front. Biosci.* **2009**, *14*, 3419–3454.

(43) Harano, Y.; Imai, T.; Kovalenko, A.; Kinoshita, M.; Hirata, F. Theoretical study for partial molar volume of amino acids and polypeptides by the three-dimensional reference interaction site model. *J. Chem. Phys.* **2001**, *114*, 9506–9511.

(44) Imai, T.; Harano, Y.; Kovalenko, A.; Hirata, F. Theoretical study for volume changes associated with the helix-coil transition of peptides. *Biopolymers* **2001**, *59*, 512–519.

(45) Perlovich, G. L.; Volkova, T. V.; Bauer-Brandl, A. Towards an understanding of the molecular mechanism of solvation of drug molecules: A thermodynamic approach by crystal lattice energy, sublimation, and solubility exemplified by paracetamol, acetanilide, and phenacetin. *J. Pharm. Sci.* **2006**, *95*, 2158–2169.

(46) Perlovich, G. L.; Volkova, T. V.; Manin, A. N.; Bauer-Brandl, A. Influence of position and size of substituents on the mechanism of partitioning: A thermodynamic study on acetaminophens, hydroxybenzoic acids, and parabens. *AAPS PharmSciTech* **2008**, *9*, 205–216.

(47) Perlovich, G. L.; Volkova, T. V.; Manin, A. N.; Bauer-Brandl, A. Extent and mechanism of solvation and partitioning of isomers of substituted benzoic acids: A thermodynamic study in the solid state and in solution. *J. Pharm. Sci.* **2008**, *97*, 3883–3896.

(48) Surov, A. O.; Szterner, P.; Zielenkiewicz, W.; Perlovich, G. L. Thermodynamic and structural study of tolfenamic acid polymorphs. *J. Pharm. Biomed. Anal.* **2009**, *50*, 831–840.

(49) Perlovich, G. L.; Surov, A. O.; Hansen, L. K.; Bauer-Brandl, A. Energetic aspects of diclofenac acid in crystal modifications and in solutions - Mechanism of solvation, partitioning and distribution. *J. Pharm. Sci.* **2007**, *96*, 1031–1042.

(50) Perlovich, G. L.; Surov, A. O.; Bauer-Brandl, A. Thermodynamic properties of flufenamic and niflumic acids - Specific and non-specific interactions in solution and in crystal lattices, mechanism of solvation, partitioning and distribution. *J. Pharm. Biomed. Anal.* **2007**, *45*, 679–687.

(51) Geballe, M. T.; Skillman, A. G.; Nicholls, A.; Guthrie, J. P.; Taylor, P. J. The SAMPL2 blind prediction challenge: introduction and overview. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 259–279.

(52) Nicholls, A.; Mobley, D. L.; Guthrie, J. P.; Chodera, J. D.; Bayly, C. I.; Cooper, M. D.; Pande, V. S. Predicting small-molecule solvation

free energies: An informal blind test for computational chemistry. *J. Med. Chem.* **2008**, *51*, 769–779.

(53) Bennaïm, A. Standard Thermodynamics Of Transfer - Uses And Misuses. *J. Phys. Chem.* **1978**, *82*, 792–803.

(54) Ben-Naim, A.; Marcus, Y. Solvation thermodynamics of non-ionic solutes. *J. Chem. Phys.* **1984**, *81*, 2016–2027.

(55) Guthrie, J. P. A Blind Challenge for Computational Solvation Free Energies: Introduction and Overview. *J. Phys. Chem. B* **2009**, *113*, 4501–4507.

(56) Lue, L.; Blankschtein, D. Liquid-state theory of hydrocarbon water-systems application to methane, ethane, and propane. *J. Phys. Chem.* **1992**, *96*, 8582–8594.

(57) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

(58) Hirata, F.; Rossky, P. J. An Extended Rism Equation For Molecular Polar Fluids. *Chem. Phys. Lett.* **1981**, *83*, 329–334.

(59) Lee, P. H.; Maggiora, G. M. Solvation Thermodynamics Of Polar-Molecules In Aqueous-Solution By The Xrism Method. *J. Phys. Chem.* **1993**, *97*, 10175–10185.

(60) Kovalenko, A.; Hirata, F. Hydration free energy of hydrophobic solutes studied by a reference interaction site model with a repulsive bridge correction and a thermodynamic perturbation method. *J. Chem. Phys.* **2000**, *113*, 2793–2805.

(61) Allen, M. P.; Tildesley, D. J., Eds. *Computer Simulation of Liquids*; Clarendon Press: Oxford, 1987.

(62) Kovalenko, A.; Hirata, F. Self-consistent description of a metal-water interface by the Kohn-Sham density functional theory and the three-dimensional reference interaction site model. *J. Chem. Phys.* **1999**, *110*, 10095–10112.

(63) Kovalenko, A.; Hirata, F. Potentials of mean force of simple ions in ambient aqueous solution. I. Three-dimensional reference interaction site model approach. *J. Chem. Phys.* **2000**, *112*, 10391–10417.

(64) Kovalenko, A. In *Molecular theory of solvation*; Hirata, F., Ed.; Kluwer Academic Publishers: Dordrecht, Netherlands, 2003; pp 169–275.

(65) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.

(66) Kovalenko, A.; Ten-No, S.; Hirata, F. Solution of three-dimensional reference interaction site model and hypernetted chain equations for simple point charge water by modified method of direct inversion in iterative subspace. *J. Comput. Chem.* **1999**, *20*, 928–936.

(67) Perkyns, J. S.; Pettitt, B. M. A Dielectrically Consistent Interaction Site Theory For Solvent Electrolyte Mixtures. *Chem. Phys. Lett.* **1992**, *190*, 626–630.

(68) Case, D. A.; et al. *Amber*, version 11; WWW page, 2010; <http://ambermd.org>.

(69) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. Amber A Package of Computer-programs For Applying Molecular Mechanics, Normal-mode Analysis, Molecular-dynamics and Free-energy Calculations To Simulate the Structural and Energetic Properties of Molecules. *Comput. Phys. Commun.* **1995**, *91*, 1–41.

(70) Wang, J. M.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247–260.

(71) Kolossvary, I.; C., G. W. Low Mode Search. An efficient, automated computational method for conformational analysis: application to cyclic and acyclic alkanes and cyclic peptides. *J. Am. Chem. Soc.* **1996**, *118*, 5011–5019.

(72) Kolossvary, I.; C., G. W. Low Mode Search Elucidated. Application to C39H80 and Flexible Docking of 9-Deazaguanine Inhibitors to PNP. *J. Comput. Chem.* **1999**, *20*, 1671–1684.

(73) *MacroModel*, version 9.8; Schrodinger, LLC: New York, NY, 2010.

(74) Sergiievskyi, V. P.; Hackbusch, W.; Fedorov, M. V. Multigrid solver for the reference interactions site model of molecular liquids theory. *J. Comput. Chem.* **2011**, *32*, 1982–1992.



(75) Marenich, A. V.; Olson, R. M.; Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. Self-consistent reaction field model for aqueous and nonaqueous solutions based on accurate polarized partial charges. *J. Chem. Theory Comput.* **2007**, *3*, 2011–2033.

(76) Olson, R. M.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Charge model 4 and intramolecular charge polarization. *J. Chem. Theory Comput.* **2007**, *3*, 2046–2054.

(77) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. SM6: A density functional theory continuum solvation model for calculating aqueous solvation free energies of neutrals, ions, and solute-water clusters. *J. Chem. Theory Comput.* **2005**, *1*, 1133–1152.

(78) *Jaguar*, version 7.7; Schrodinger, LLC: New York, NY, 2010.

(79) Frisch, M. J.; et al. *Gaussian 03*; Gaussian, Inc.: Wallingford, CT, 2004.

(80) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Performance of SM6, SM8, and SMD on the SAMPL1 Test Set for the Prediction of Small-Molecule Solvation Free Energies. *J. Phys. Chem. B* **2009**, *113*, 4538–4543.

(81) Nicholls, A.; Wlodek, S.; Grant, J. SAMPL2 and continuum modeling. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 293–306.

(82) Meunier, A.; Truchon, J.-F. Predictions of hydration free energies from continuum solvent with solute polarizable models: the SAMPL2 blind challenge. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 361–372.

(83) Imai, T.; Kovalenko, A.; Hirata, F. Solvation thermodynamics of protein studied by the 3D-RISM theory. *Chem. Phys. Lett.* **2004**, *395*, 1–6.

(84) Stumpe, M. C.; Blinov, N.; Wishart, D.; Kovalenko, A.; Pande, V. S. Calculation of Local Water Densities in Biological Systems: A Comparison of Molecular Dynamics Simulations and the 3D-RISM-KH Molecular Theory of Solvation. *J. Phys. Chem. B* **2011**, *115*, 319–328.