

Implementation and Optimization of the Embedded Cluster Reference Interaction Site Model with Atomic Charges

Ádám Ganyecz* and Mihály Kállay



Cite This: *J. Phys. Chem. A* 2022, 126, 2417–2429



Read Online

ACCESS |



Metrics & More

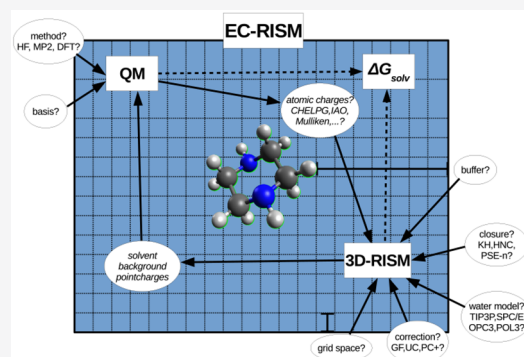


Article Recommendations



Supporting Information

ABSTRACT: In this work, we implemented the embedded cluster reference interaction site model (EC-RISM) originally developed by Kloss, Heil, and Kast (*J. Phys. Chem. B* 2008, 112, 4337–4343). This method combines quantum mechanical calculations with the 3D reference interaction site model (3D-RISM). Numerous options, such as buffer, grid space, basis set, charge model, water model, closure relation, and so forth, were investigated to find the best settings. Additionally, the small point charges, which are derived from the solvent distribution from the 3D-RISM solution to represent the solvent in the QM calculation, were neglected to reduce the overhead without the loss of accuracy. On the MNSOL[a], MNSOL, and FreeSolv databases, our implemented and optimized method provides solvation free energies in water with 5.70, 6.32, and 6.44 kJ/mol root-mean-square deviations, respectively, but with different settings, 5.22, 6.08, and 6.63 kJ/mol can also be achieved. Only solvent models containing fitting parameters, like COSMO-RS and EC-RISM with universal correction and directly used electrostatic potential, perform better than our EC-RISM implementation with atomic charges.



INTRODUCTION

Most processes take place in solution therefore the solvent effect has to be considered during theoretical approaches if appropriate description of the system is desired. Not only reactions, but also several physicochemical properties are connected to solvation, such as pK_a , dissociation and complexation constants, solvation free energies, $\log D/\log P$, solubilities, and so forth.¹

One way to determine these properties is to use quantitative structure–property relationships (QSPR), which are usually empirically parametrized functions that relate a target property to a set of molecular descriptors calculable from a simple computational representation of the molecule.² Even though it is a cheap method, this approach lacks the ability to predict accurate values for species which differ considerably from the training data and does not provide insight into the solvation processes.

The other way to predict the physicochemical properties of solvation is to use a solvation model together with the appropriate theoretical computational approach. Solvent molecules can then be treated explicitly which can give a realistic picture of the solvation processes. Because a large number of solvent molecules are needed, this approach is used with molecular dynamics (MD) or Monte Carlo methods (MC), and it is still computationally expensive.^{3,4}

In contrast, implicit solvent models treat the solvent as a homogeneously polarizable medium. This greatly reduces the computational costs and enables the usage of quantum mechanical methods in the solvent phase. There are numerous

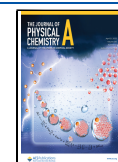
implicit solvent models [polarizable continuum model (PCM),⁵ integral equation formalism PCM (IEFPCM),^{6,7} conductor-like PCM (CPCM),⁸ conductor-like screening model (COSMO),⁹ COSMO for realistic solvation (COSMO-RS),¹⁰ direct COSMO-RS (DCOSMO-RS),¹¹ solvation model based on charge density (SMD),¹² and so forth], which are discussed elsewhere in detail.¹³ However, these methods lack the ability to describe specific interactions between the solute and solvent.

Another class of solvation approaches is based on integral equation theory (IET). It can provide information about the solvent density around the solute without the need of simulating thousands of solvent molecules, leading to a cost comparable to implicit solvent models. Due to these features, remarkable development has been made in recent years regarding methods exploiting IET, like the reference interaction site model (RISM)^{14–16} or molecular density functional theory (MDFT).^{17–22} About its theory and the various methods, the interested reader is referred to the review of Ratkova et al.¹ Currently, the most successful formulation of IET is the three-dimensional RISM (3D-RISM) approach.^{15,16}

Received: September 7, 2021

Revised: March 22, 2022

Published: April 8, 2022



However, there are still challenges concerning the implementation of this model. First, one needs to find the so-called bridge function to solve the RISM equations. There exist no exact functions, but several applicable closure functions have been proposed for that purpose, like the hypernetted-chain (HNC),^{23,24} the Kovalenko–Hirata (KH),^{25,26} and the partial series expansion of order n (PSE- n) of HNC closure²⁷ approaches. Second, the solvation free energies supplied by the theory are far from the experimental ones. This problem is usually treated with a correction after the calculation based on the partial molar volume of the solute, which is already available from the 3D-RISM solution, either by an empirical approach like the universal correction (UC)²⁸ or based on theoretical considerations like the pressure correction (PC and PC+) approximation.^{29–31}

Because of the attractive features of IET, there were several attempts to combine it with the QM treatment of the solute. Ten-no et al. coupled the Hartree–Fock (HF) self-consistent field (SCF) method with the RISM theory (RISM-SCF),³² which was later further improved by Yokogawa et al. with the RISM-SCF-SEDD approach.³³ Extensions of these methods to 3D-RISM have also been developed.^{34–36} All of these methods solve the RISM equations in every SCF iteration step with partial charges or electron density from the current SCF step providing the electrostatic potential on solute molecules for the next SCF step. Kloss, Heil, and Kast chose another approach for the coupling of QM and RISM called embedded cluster reference interaction site model (EC-RISM).³⁷ In this case, the QM solution provides partial charges of the solute to solve the 3D-RISM equations, which gives background point charges describing the solvent for the next cycle of QM calculation. The main advantages are that the QM code and the RISM solver (like AmberTools³⁸) can be easily combined, and the self-consistent approach converge well. In the past decade, EC-RISM has been further developed to use exact electrostatics instead of atomic charges, and used successfully in the SAMPL challenges to predict solvation free energies, and pK_a , $\log D$, and $\log P$ values.^{39–43}

Here, we report the implementation of the EC-RISM approach with atomic charges of Kloss, Heil, and Kast³⁷ and optimize the various RISM and QM settings regarding solvation free energies. We compare our results to other solvation models using the MNSOL⁴⁴ and FreeSolv databases.^{4,45}

METHODS

Theory. From IET to 3D-RISM. The RISM theory was discussed in detail elsewhere,^{24,46,47} here, we just provide a brief outline. The starting point of IET is the Ornstein–Zernike (OZ) equation,⁴⁸ which defines the total correlation function $h(r)$ between a pair of spherical particles:

$$h(r_{12}) = c(r_{12}) + \rho \int c(r_{13})h(r_{32}) dr_3 \quad (1)$$

where ρ is the density of the homogeneous isotropic fluid, $c(r)$ is the direct correlation function. The pair density distribution function (also known as the pair correlation or pair distribution function) $g(r)$ can be written as $h(r) + 1$. According to the OZ equation, the total correlation between two particles is the sum of the direct correlation function and an indirect correlation function [$\gamma(r_{12}) = h(r_{12}) - c(r_{12})$], which consists of an infinite series of direct correlations.

The molecular OZ (MOZ) equation is a generalization of the OZ equation to nonspherical molecules.^{24,49} Here, the correlation functions are also dependent on the orientations of the molecules, leading to 6 dimensional equations, making the usage of MOZ nonpractical.

On the basis of the work of Chandler and Anderson,¹⁴ RISM methods have emerged that reduce the high dimensionality of the MOZ equations. In the one-dimensional RISM (1D-RISM) approach, one-dimensional integral equations and intermolecular spherically symmetric site–site correlation functions are used. The spherical symmetry results in correlation functions that only depend on the r distance between sites as

$$h_{s\alpha}(r) = \sum_{s'=1}^{M_{\text{solute}}} \sum_{\xi=1}^{N_{\text{solvent}}} \int_{R^3} \int_{R^3} \omega_{s\xi}(|\mathbf{r}_1 - \mathbf{r}'|) \times c_{s'\xi}(|\mathbf{r}' - \mathbf{r}''|) \chi_{\xi\alpha}(|\mathbf{r}'' - \mathbf{r}_2|) d\mathbf{r}' d\mathbf{r}'' \quad (2)$$

where s refers to the solute sites, whereas α and ξ to the solvent sites. The structures of the molecules are described by the intramolecular correlation function, $\omega(r)$, while $\chi_{\xi\alpha}(r)$ is the reduced bulk solvent susceptibility function, and M_{solute} and N_{solvent} are the number solute and solvent sites. There are several different realizations of 1D-RISM, such as the extended RISM (XRISM),⁵⁰ ARISM⁵¹ (“A” refers to a scaling coefficient), and dielectrically consistent RISM (DRISM).⁵² The work of Beglov and Roux, and Kovalenko and Hirata introduced the three-dimensional extension of RISM, 3D-RISM.^{15,16,53} In this case, the MOZ equation is replaced by a set of 3D integral equations through partial integration over the orientational coordinates. This leads to the following equations with intermolecular solvent site–solute total correlation functions, $h_{\alpha}(r)$, and direct correlation functions, $c_{\alpha}(r)$, which are more feasible:

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

The reduced solvent susceptibility function can be calculated from a 1D-RISM calculation by

$$\chi_{\xi\alpha}(r) = \omega_{\xi\alpha}^{\text{solv}}(r) + \rho h_{\xi\alpha}^{\text{solv}}(r) \quad (4)$$

The solvent susceptibility function can be obtained by multiplying χ with ρ .

As mentioned above, the solution of OZ and RISM requires a so-called closure relation because we have more unknown functions than equations:^{24,46,49}

$$h(r) + 1 = \exp[-\beta u(r) + \gamma(r) + B(r)] \quad (5)$$

where $B(r)$ stands for the bridge function, which is a functional of the indirect correlation function, $\beta = 1/k_B T$, k_B denotes the Boltzmann constant, and T is the absolute temperature. $u(r)$ stands for the pair interaction potential, which is usually described by the Coulomb and Lennard–Jones interactions in the RISM formalism. Unfortunately, the exact expression of the bridge function is unknown, so approximations are needed.

The simplest closure is HNC, where the $B(r)$ term is simply ignored [$B(r) = 0$],⁵⁴ however, this closure often leads to convergence issues. Kovalenko and Hirata introduced a partially linearized version of the HNC closure (PLHNC):²⁶

$$h_\alpha(\mathbf{r}) = \begin{cases} \exp[-\beta u_\alpha(\mathbf{r}) + \gamma_\alpha(\mathbf{r})] - 1 & \text{if } -\beta u_\alpha(\mathbf{r}) + \gamma_\alpha(\mathbf{r}) \leq C \\ -\beta u_\alpha(\mathbf{r}) + \gamma_\alpha(\mathbf{r}) + \exp(C) - C - 1 & \text{if } -\beta u_\alpha(\mathbf{r}) + \gamma_\alpha(\mathbf{r}) > C \end{cases} \quad (6)$$

If $C = 0$, the PLHNC closure becomes the Kovalenko–Hirata (KH) closure. Kast and Kloss used partial series expansion of order n (PSE- n) of the HNC closure to tackle the convergence problems of HNC:²⁷

$$h_\alpha(\mathbf{r}) = \begin{cases} \exp[-\beta u_\alpha(\mathbf{r}) + \gamma_\alpha(\mathbf{r})] - 1 & \text{if } -\beta u_\alpha(\mathbf{r}) + \gamma_\alpha(\mathbf{r}) \leq 0 \\ \sum_{i=0}^n [-\beta u_\alpha(\mathbf{r}) + \gamma_\alpha(\mathbf{r})]^i / i! - 1 & \text{if } -\beta u_\alpha(\mathbf{r}) + \gamma_\alpha(\mathbf{r}) > 0 \end{cases} \quad (7)$$

Relying on the correlation functions from the solution of 3D-RISM equations, we can determine the solvation free energy (ΔG_{solv}) using Kirkwood's equation.⁵⁵ The corresponding free energy functionals for the above-mentioned closures are

$$\Delta G_{\text{solv}}^{\text{HNC}} = k_{\text{B}}T \sum_{\alpha=1}^{N_{\text{solvent}}} \rho_\alpha \int_{\mathbb{R}^3} \left[\frac{1}{2} h_\alpha^2(\mathbf{r}) - c_\alpha(\mathbf{r}) - \frac{1}{2} c_\alpha(\mathbf{r}) h_\alpha(\mathbf{r}) \right] d\mathbf{r} \quad (8)$$

$$\Delta G_{\text{solv}}^{\text{PSE-}n} = \Delta G_{\text{solv}}^{\text{HNC}} - k_{\text{B}}T \sum_{\alpha=1}^{N_{\text{solvent}}} \rho_\alpha \int_{\mathbb{R}^3} \left[\frac{\Theta(-\beta u_\alpha(\mathbf{r}) + h_\alpha(\mathbf{r}) - c_\alpha(\mathbf{r}))(-\beta u_\alpha(\mathbf{r}) + h_\alpha(\mathbf{r}) - c_\alpha(\mathbf{r}))^{n+1}}{(n+1)!} \right] d\mathbf{r} \quad (9)$$

$$\Delta G_{\text{solv}}^{\text{KH}} = k_{\text{B}}T \sum_{\alpha=1}^{N_{\text{solvent}}} \rho_\alpha \int_{\mathbb{R}^3} \left[\frac{1}{2} h_\alpha^2(\mathbf{r}) \Theta(-h_\alpha(\mathbf{r})) - c_\alpha(\mathbf{r}) - \frac{1}{2} c_\alpha(\mathbf{r}) h_\alpha(\mathbf{r}) \right] d\mathbf{r} \quad (10)$$

where ρ_α is the number density of solvent site α , and Θ denotes the Heaviside step function:

$$\Theta(x) = \begin{cases} 1 & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases} \quad (11)$$

Unfortunately, the accuracy of solvation free energies obtained by the 3D-RISM method is not satisfactory. One of the first models to solve this problem assumes Gaussian fluctuations of the solvent molecules (GF) and evaluates the solvation free energy as^{56,57}

$$\Delta G_{\text{solv}}^{\text{GF}} = k_{\text{B}}T \sum_{\alpha=1}^{N_{\text{solvent}}} \rho_\alpha \int_{\mathbb{R}^3} \left[-c_\alpha(\mathbf{r}) - \frac{1}{2} c_\alpha(\mathbf{r}) h_\alpha(\mathbf{r}) \right] d\mathbf{r} \quad (12)$$

One of the most successful correction approaches utilizing this approximation is the universal correction (UC).²⁸ This model is based on the observation that the solvation free energy errors are linearly dependent on the dimensionless partial molar volume (DPMV, $\rho \bar{V}$) of the solvents,⁵⁸ which can also be determined from 3D-RISM. Therefore, a linear fit on experimental data can produce the fitting parameters (a_1^{GF} , a_0^{GF}), and the solvation free energy is augmented by a correction as

$$\Delta G_{\text{solv}}^{\text{UC}} = \Delta G_{\text{solv}}^{\text{GF}} + a_1^{\text{GF}}(\rho \bar{V}) + a_0^{\text{GF}} \quad (13)$$

This expression can be used with any free energy functional not only with GF. Theoretically more established models are the pressure correction (PC) and advanced pressure correction (PC+),^{29–31} which employ the following expressions:

$$\Delta G_{\text{solv}}^{\text{PC}} = \Delta G_{\text{solv}}^{\text{3D-RISM}} - P^{\text{RISM}} \bar{V} \quad (14)$$

$$\Delta G_{\text{solv}}^{\text{PC+}} = \Delta G_{\text{solv}}^{\text{3D-RISM}} - P^{\text{RISM}} \bar{V} + \rho k_{\text{B}}T \bar{V} \quad (15)$$

where P^{RISM} is the pressure calculated from the solvent–solvent direct correlation function as

$$P^{\text{RISM}} = \frac{N_{\text{solvent}} + 1}{2} \rho k_{\text{B}}T - \frac{k_{\text{B}}T}{2} \sum_{\alpha=1}^{N_{\text{solvent}}} \sum_{\beta=1}^{N_{\text{solvent}}} \rho_\alpha \rho_\beta \hat{c}_{\alpha\beta}(k=0) \quad (16)$$

where $\hat{c}_{\alpha\beta}(k=0)$ is the Fourier transformed direct correlation function at $k=0$.

EC-RISM. RISM as a solvation model can be combined with QM and MM methods. In this work, we focus on the EC-RISM method of Kast and co-workers,^{37,59} which was adopted using the MRCC quantum chemistry program⁶⁰ for the QM calculations and the AmberTools MD package³⁸ for the 3D-RISM calculations. Here, we briefly present the theory behind EC-RISM, which is already explained in detail in the original EC-RISM articles.^{37,59}

In this framework, the solvation free energy from 3D-RISM is corrected with the electronic energy change of the solute due to the solvent ($\Delta E^{\text{QM}} = E_{\text{solv}}^{\text{QM}} - E_{\text{gas}}^{\text{QM}}$):

$$\Delta G_{\text{solv}}^{\text{EC-RISM}} = \Delta E^{\text{QM}} + \Delta G_{\text{solv}}^{\text{3D-RISM}} \quad (17)$$

In this approach, where the solvent is represented as background point charges in the QM calculations, the Hamiltonian of the solute can be written as

$$\hat{H}_{\text{tot}} = \hat{H}_1 + \hat{H}_2 \quad (18)$$

$$\hat{H}_1 = \hat{H}_{\text{ne}} + \hat{H}_{\text{ee}} + \hat{H}_{\text{nn}} \quad (19)$$

$$\hat{H}_2 = \hat{H}_{\text{nq}} + \hat{H}_{\text{eq}} + \hat{H}_{\text{qq}} \quad (20)$$

where n , e , and q refer to the interaction between the nuclei, electrons, and point charges as appropriate.

The energy of the solute, whose wave function (ψ_{tot}) is affected by the background charges, can be written as

$$E_{\text{solv}}^{\text{QM}} = \langle \psi_{\text{tot}} | \hat{H}_1 | \psi_{\text{tot}} \rangle = E_{\text{tot}}^{\text{QM}} - E_2^{\text{QM}} \quad (21)$$

Therefore, we have to subtract all point charge related energy terms, collected in E_2^{QM} as

$$E_2^{\text{QM}} = \langle \psi_{\text{tot}} | \hat{H}_{\text{nq}} + \hat{H}_{\text{eq}} | \psi_{\text{tot}} \rangle + \langle \psi_{\text{tot}} | \hat{H}_{\text{qq}} | \psi_{\text{tot}} \rangle = E_{\text{q}}^{\text{QM}} + E_{\text{qq}}^{\text{QM}} \quad (22)$$

from the total energy. Here, $E_{\text{qq}}^{\text{QM}}$ is the self-energy of the background point charges, and its calculation can be omitted due to the fact that it is also included in $E_{\text{tot}}^{\text{QM}}$, and its computation can be demanding when dealing with hundreds of thousands or millions of point charges. E_{q}^{QM} is the electrostatic interaction energy of the point charges with the solute, which is already taken into account during the 3D-RISM calculations. It can be described as the interaction of the electrostatic potential of the solute (φ) and the charge density ρ_q at spatial points \mathbf{r} , which can be simplified to the summation of the products of the background charges [$q(\mathbf{r}_i)$] and the electrostatic potential at each grid point:

$$E_{\text{q}}^{\text{QM}} = \int \rho_q(\mathbf{r}) \varphi(\mathbf{r}) d\mathbf{r} \approx \sum_i q(\mathbf{r}_i) \varphi(\mathbf{r}_i) \quad (23)$$

The required background point charges can be derived from the distribution functions g from the 3D-RISM solution by taking the sum of the products of its value [$g_{\gamma}(\mathbf{r}_i)$] and the charge of the corresponding site (q_{γ}):

$$q(\mathbf{r}_i) = \sum_{\gamma} g_{\gamma}(\mathbf{r}_i) q_{\gamma} \quad (24)$$

The 3D-RISM calculation requires the electrostatic potential for the Coulombic interaction of the interaction potential (u), which is represented with atomic charges based on the QM calculation. The rest of the input for 3D-RISM is independent of the QM part.

The general workflow of the EC-RISM algorithm is presented in Figure 1. The procedure is initialized by a QM

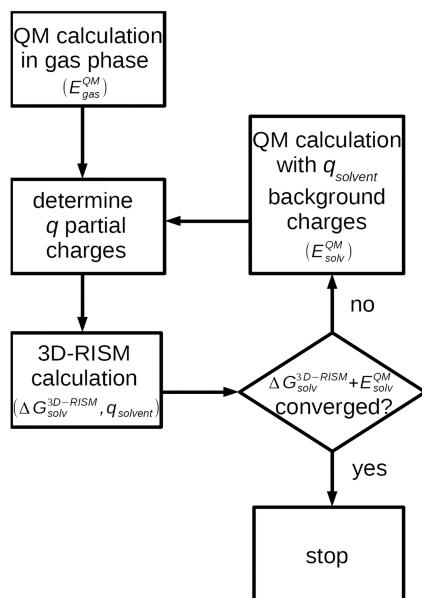


Figure 1. Flowchart of our EC-RISM workflow.

calculation in gas phase. Then, the following steps are repeated until self-consistency, that is, the energy change in QM and RISM is below 0.01 kJ/mol: (i) the solute atomic charges are determined based on a QM calculation, (ii) a 3D-RISM calculation is carried out using the QM atomic charges, and (iii) a QM calculation is performed utilizing the solvent charge distribution obtained from the 3D-RISM solution as background point charges representing the solvent. We note that, instead of using atomic charges to represent the electrostatic potential in the 3D-RISM calculations, the electrostatic potential could also be used directly.^{39–41} The algorithm has been implemented in MRCC using the AmberTools package for the 3D-RISM calculations. Initial work was aided by the python script of Misin et al.^{30,61}

EC-RISM has been used in the SAMPL challenges to predict solvation free energies, and pK_a , $\log D$, and $\log P$ values.^{41–43,59,62} Besides that, the EC-RISM approach has been utilized successfully to study the effect of pressure on various properties,⁶³ such as the pressure response of the dipole moment of trimethylamine-*N*-oxide (TMAO) in aqueous solution,⁶⁴ the pressure dependence of NMR chemical shifts for *N*-methylacetamide,⁶⁵ to understand the pressure-induced blue-shift of IR bands of TMAO,⁶⁶ and to develop a force field for urea in aqueous solutions under high pressure.⁶⁷

Computational Details. As a reference for experimental solvation free energies, the MNSOL⁴⁴ and the FreeSolv⁴ databases were used in this work. We employed the MNSOL[a] data set, which is a subset of the whole MNSOL database and the most commonly used data set for testing and optimizing solvation models. Conformers were generated with OPENBABEL⁶⁸ and optimized at the B3LYP/def2-TZVP level of theory with MRCC in gas phase without relaxing in solution phase. During the optimization of the EC-RISM settings, only the most stable conformers were used with the MNSOL[a] data set, while for the final analysis conformers with energies of at most 10 kJ/mol higher than the most stable one were considered with the MNSOL and FreeSolv databases. The conformationally averaged results were obtained using Boltzmann averaging.

For the QM part of the EC-RISM calculations, various settings were considered. As QM methods, HF, second-order Møller-Plesset (MP2), and several DFT functionals were utilized applying the density-fitting (DF) approximation. Concerning the atomic orbital basis sets, the correlation consistent (aug-)cc-pVXZ^{69–71} and def2⁷² bases were tested. Atomic charges are derived from the density of the method used, for example, MP2 density is applied to generate atomic charges if MP2 calculations are performed.

Water susceptibility functions were prepared with the IDRISM program of AmberTools³⁸ invoking the DRISM theory.⁵² The density was set to 55.343 mol/dm³, the dielectric constant to 78.375, and the grid spacing to 0.025 Å. The water models utilized and their parameters are collected in Table 1. The original SPC/E,⁷³ TIP3P,⁷⁴ OPC3⁷⁵ and POL3⁷⁶ water models are not sufficient, because the Lennard-Jones parameters of H are needed to avoid the convergence issues of 1D-RISM. Pettitt and Rossky used $\sigma_H = 0.4$ Å and $\epsilon_H = 0.192464$ kJ/mol,⁷⁷ while Hirata and co-workers applied $\sigma_H = 1.0$ Å and $\epsilon_H = 0.2282372$ kJ/mol.⁷⁸ Kast et al. employed $\sigma_H = 1.0$ Å and $\epsilon_H = 0.234304$ kJ/mol for their modified SPC/E water model⁵⁹ (mSPC/E) based on the work of Maw et al.⁷⁹ In this work, we opted for the approach of Luchko et al.,⁸⁰ who

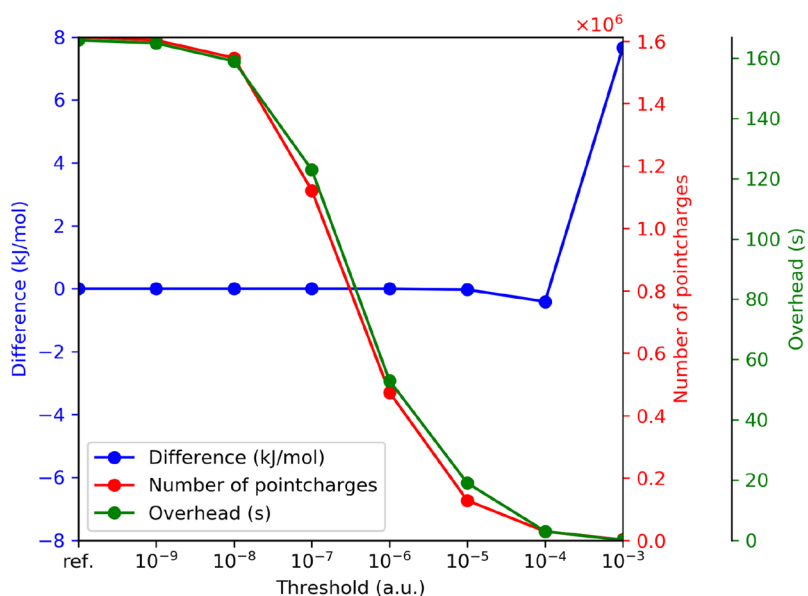
Table 1. Parameters of the Water Models Employed in This Work

	cSPC/E	cTIP3P	cOPC3	cPOL3	mSPC/E
r_{OH} (Å)	1.0000	0.9572	1.0000	0.9789	1.0000
θ_{HOH} (deg)	109.47	104.52	109.47	109.47	109.47
q_O (e)	−0.8476	−0.8340	−0.7300	−0.8952	−0.8476
q_H (e)	0.4238	0.4170	0.3650	0.4476	0.4238
σ_O (Å)	3.1658	3.1507	3.2037	3.1743	3.1658
σ_H (Å)	1.1658 ^a	1.2363 ^a	1.2037 ^a	1.2165 ^a	1.0000
ϵ_O (kJ/mol)	0.64978	0.63597	0.65270	0.68369	0.64978
ϵ_H (kJ/mol)	0.064978 ^b	0.063597 ^b	0.065270 ^b	0.068369 ^b	0.234304

^aCalculated using $\frac{\sigma_H}{2} = \frac{\sigma_O}{2} - r_{OH}$.⁸⁰ ^bCalculated using $\epsilon_H = 0.1\epsilon_O$.⁸⁰

Table 2. Comparison of 3D-RISM Solvation Free Energies (in kJ/mol) and Runtimes (in s) with Different Buffer and Grid Space for Piperazine with PSE-3 Closure, cSPC/E Water Model, and PC+ Correction

grid space (Å)	0.1		0.2		0.3		0.4		0.5	
buffer (Å)	ΔG_{solv}	time	ΔG_{solv}	time	ΔG_{solv}	time	ΔG_{solv}	time	ΔG_{solv}	time
5	-36.01	41.2	-35.96	4.2	-35.88	1.3	-35.81	0.5	-35.40	0.3
10	-36.11	175.1	-36.07	21.5	-36.09	6.3	-36.07	2.1	-35.71	1.2
15	-36.11	494.5	-36.07	64.7	-36.10	19.8	-36.07	6.2	-35.72	3.2
20	-36.12	959.8	-36.07	123.2	-36.11	38.9	-36.08	13.2	-35.73	7.0
25	-36.12	1960.6	-36.07	240.2	-36.11	78.9	-36.08	26.5	-35.73	14.1
30	-36.13	3112.4	-36.07	389.1	-36.11	120.0	-36.09	44.7	-35.73	26.8
35	-36.13	4758.8	-36.07	608.3	-36.11	173.9	-36.09	70.7	-35.73	35.4
40			-36.06	863.0	-36.11	276.9	-36.09	104.1	-35.73	62.5
45			-36.03	1235.9	-36.11	356.5	-36.09	135.5	-35.72	73.1
50			-36.02	1774.7	-36.08	468.8	-36.06	207.6	-35.71	94.7

**Figure 2.** Effect of neglecting background point charges below a threshold on the energy and the runtime. The reference is a calculation where there are no point charges dropped from the RISM calculation with applying a 15 Å buffer, 0.3 Å grid space, cSPC/E water, PC+ correction, and PSE-3 closure. For comparison, the DF-MP2/def2-TZVPPD gas-phase calculation of piperazine requires 101 s on an Intel(R) Xeon(R) CPU E5-1650 v2 @ 3.50 GHz processor with 12 cores.

proposed the $\frac{\sigma_{\text{H}}}{2} = \frac{\sigma_{\text{O}}}{2} - r_{\text{OH}}$ and $\epsilon_{\text{H}} = 0$. $1\epsilon_{\text{O}}$ rules. Their modified SPC/E and TIP3P model is referred to as coincident SPC/E (cSPC/E) and coincident TIP3P (cTIP3P). The modified OPC3 and POL3 models are produced following the same approach and will be called cOPC3 and cPOL3. Table S1 of the Supporting Information (SI) presents the compressibilities obtained, and the PC and PC+ correction factors employed in this work.

For the 3D-RISM calculations, the RISM3D.SNGLPNT program of AmberTools was utilized with various closure, buffer, grid space, and water models. The GAFF2 force field was used in every calculation. During the work, the temperature was set to 298.15 K where it was necessary. The partial molar volume, which is needed for some corrections, is calculated with the following formula in RISM3D.SNGLPNT:

$$\hat{V} = k_{\text{B}}T\chi \left(1 - \rho \sum_{\alpha=1}^{N_{\text{solvent}}} \int_{\mathbb{R}^3} c_{\alpha}(\mathbf{r}) d\mathbf{r} \right) \quad (25)$$

The parameters for Si were adopted from the work of Dong et al.: a van der Waals radius of 1.778 Å and a well-depth of 0.015 kcal/mol were applied.⁸¹

RESULTS AND DISCUSSION

Optimization of the Settings for EC-RISM. As the first step, we optimized those settings that can be arbitrarily improved. These are the buffer and grid space for the RISM calculations and the basis set for the QM calculations. As test molecules, we chose a medium-sized druglike species, piperazine, and a smaller more polar one, NH₃. The results for piperazine are discussed here, whereas those for NH₃ are presented in Table S2. In general, there is no significant difference between the conclusions which can be drawn for the two test systems. Other settings, like the used charge model, water model, closure, and correction, cannot be optimized without references because, in contrast to the aforementioned settings, there is no theoretical limit that could provide a reference. Therefore, these are optimized by taking the MNSOL[a] data set, and we compare the results to experimental references.

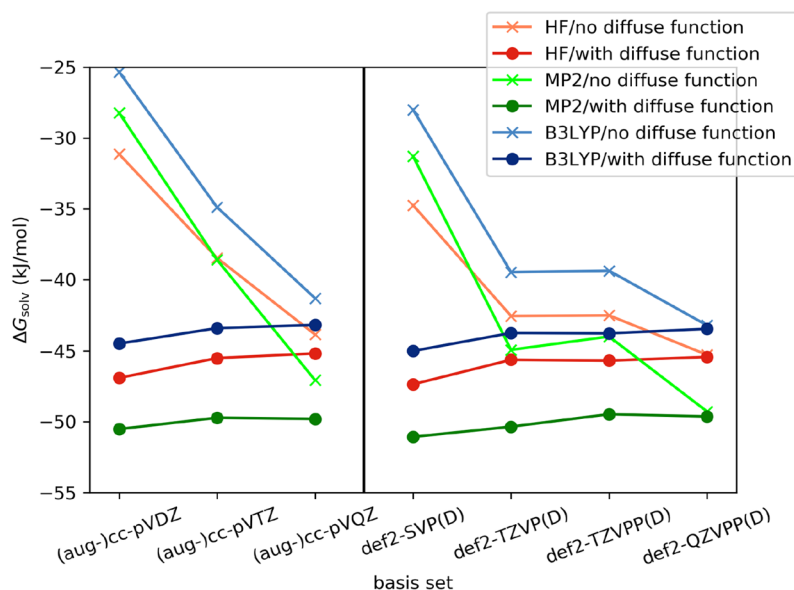


Figure 3. Effect of the basis set on the EC-RISM solvation free energies of piperazine using PSE-3 closure, PC+ correction, cSPC/E water, 15 Å buffer, and 0.3 Å grid space with CHELPG charges.

Buffer and Grid Space. The buffer is the distance between the solute and the side of the simulation box, therefore it determines the dimensions of the simulation box. The grid space is practically the distance between the grid points. Larger buffer and smaller grid space obviously improve the accuracy of RISM calculations but increase the cost. To optimize them, piperazine was used as a test molecule using the PSE-3 closure with the PC+ correction and the cSPC/E water model using buffers of 5, 10, 15, ..., 50 Å and grid spaces of 0.1, 0.2, 0.3, 0.4, and 0.5 Å. The results can be seen in Table 2. It can be observed that, with a buffer of 5 Å, the RISM solvation free energy differs at least with 0.1 kJ/mol from that obtained with larger buffers. Out of the tested grid spaces, only the 0.5 Å results differ from the smaller buffers with at least 0.1 kJ/mol (0.3–0.4 kJ/mol). However, using a more polar molecule, such as NH₃, the results with grid spaces of 0.4 Å differ from the smaller ones with 0.08–0.11 kJ/mol. Therefore, we chose a buffer of 15 Å and a grid space of 0.3 Å for the RISM calculations as a compromise between accuracy and cost.

Cost Reduction by Dropping Small Charges. The treatment of the millions of point charges generated during the 3D-RISM calculation can be cumbersome. However, most of them are too small and/or too far from the molecule to have an impactful effect on the QM results. Therefore, we can drop these insignificant point charges to speed up the QM calculations. Figure 2 shows the results for the first QM calculations of the EC-RISM cycle by dropping point charges below a certain threshold from the first RISM calculation with a buffer of 15 Å, grid space of 0.3 Å, cSPC/E water, PC+ correction, and PSE-3 closure. It can be seen that charges smaller than 10⁻⁶ a.u. can be neglected without any loss of accuracy, which results in dropping 70.6% of all point charges in this setting. Also, the overhead due to the background point charges is proportional to the number of point charges. Even more point charges can be dismissed with thresholds of 10⁻⁵ or 10⁻⁴ a.u., leading to a noticeable but acceptable difference of a few hundredths or tenths of a kJ/mol. If a larger or denser grid is used for the RISM calculation, then an even bigger speedup can be achieved with this technique. With the more polar NH₃,

similar results are obtained. From now on, we use the conservative threshold of 10⁻⁶ a.u. for dropping insignificant charges.

Basis Set Dependence. The size of the basis sets can also be arbitrarily increased until we reach the basis set limit, so it is clear that this setting should also be investigated. In Figure 3, we compared the EC-RISM solvation free energies of piperazine using the previously determined settings along with the PSE-3 closure, cSPC/E water, and PC+ correction, with various basis sets of double- (DZ), triple- (TZ), and quadruple- ζ (QZ) quality at the HF, MP2 and B3LYP levels of theory. For the DZ and TZ basis sets, we can see a clear difference between the results of bases with and without diffuse functions. The most notable difference, 22.3 kJ/mol, can be observed between the MP2/cc-pVDZ and MP2/aug-cc-pVDZ results. If diffuse functions are used, the solvation free energies become almost independent of the size of the basis set. However, without diffuse functions, the EC-RISM results are heavily dependent on the basis set size, and only QZ-quality bases can approach the performance of basis sets with diffuse functions. Therefore, diffuse functions are necessary for accurate results. Even DZ-quality basis sets with diffuse functions can provide acceptable results, but the TZ-quality bases sets are preferred. Again, for the more polar NH₃, the conclusions are similar, see Table S2. For further calculations in this work, the def2-TZVPPD basis set is employed as this basis provides almost converged solvation energies and is somewhat more economical than the aug-cc-pVTZ basis, which is of similar quality.

Charge Models and Corrections. Here, the effect of various charge models (CHELPG,⁸² Merz-Kollman,⁸³ IAO,⁸⁴ Mulliken,⁸⁵ Löwdin⁸⁶) and 3D-RISM corrections (PC, PC+, Gaussian fluctuation, UC) on the quality of the computed solvation free energies is tested. The charge model is needed to determine the Coulomb interaction of the solute and the solvent for the RISM3D.SNGLPNT program. As mentioned earlier, the corrections are essential to get solvation free energies which are comparable to the experimental ones.

Table 3 shows the RMSDs of the EC-RISM solvation free energies against the MNSOL[a] reference values with the

Table 3. Effect of Various Charges and Corrections on the RMSDs of Solvation Free Energies on the MNSOL[a] Dataset (Values in kJ/mol)^a

corrections	CHELPG	MK	IAO	Mulliken ^b	Löwdin ^c
PC	15.08	17.05	19.28	34.00	45.19
PC+	5.75	6.40	10.76	26.56	41.68
UC	5.41	5.16	9.92	20.21	41.30
GF	21.66	21.27	19.73	29.88	88.55

^aValues in italic are only informative but not directly comparable to non-italic ones. ^b54 out of the 273 species are not converged. ^c32 out of the 273 species are not converged.

different charge models and corrections. The QM part was calculated at DF-MP2/def2-TZVPPD level of theory, while the PSE-3 closure with the cTIP3P water model was applied in the RISM calculations. With the Löwdin and Mulliken charges, respectively, 54 and 32 out of 273 species not converged, therefore they are not recommended. Also the accuracy of the converged ones are way worse than with the other charges with the RMSDs being above 20 kJ/mol. In line with the previous experience,⁸⁷ the electrostatically fitted charges, that is, CHELPG and MK, perform better than the other charge models studied. Out of the four corrections, the theoretically derived PC+ and the fitted UC perform similarly, but obviously UC is slightly better. Consequently, as a charge model, CHELPG and MK are recommended, while PC+ and the universal correction are preferred to determine solvation free energies. These setting together provide RMSDs between 5 and 7 kJ/mol. We prefer PC+ because it does not require the determination of fitting parameters for different settings, even though it can be somewhat less accurate.

The use of atomic charge models can be bypassed by employing the solute electrostatic potential computed by the QM approach directly in the 3D-RISM calculation as it was suggested by Frach and Kast,⁴⁰ but we have not yet implemented this approximation.

Hereafter, CHELPG charges and the PC+ correction will be utilized.

Water Models and Closures. The closure function and the water/solvent model are other settings in RISM calculations where we have several options. Figure 4 displays our results on the MNSOL[a] data set with different closure functions (KH, PSE-2, PSE-3, PSE-4, HNC) and various water models (cSPC/E, cTIP3P, cOPC3, cPOL3). We experienced convergence issues with the PSE-4 and HNC closures, which were also discussed in the literature.¹ The use of PSE-2 and PSE-3 is a good compromise between accuracy and convergence performance, while the KH closure provides results with a larger RMSD than the other closures, around 7–10 kJ/mol. The water models perform similarly yielding RMSDs of around 5–7 kJ/mol with the PSE-2 and PSE-3 closures. The best combinations are PSE-3/cTIP3P and PSE-2/cPOL3. Another observation is that the standard deviation of the differences is rather similar across the board, 5.3–5.8 kJ/mol, and these settings just shift the solvation free energies but do not change them fundamentally. From KH to HNC, the solvation free energy increases, while the increasing order for water models is cOPC3 < cSPC/E < cTIP3P < cPOL3. More detailed statistics and results can be found in Table S3.

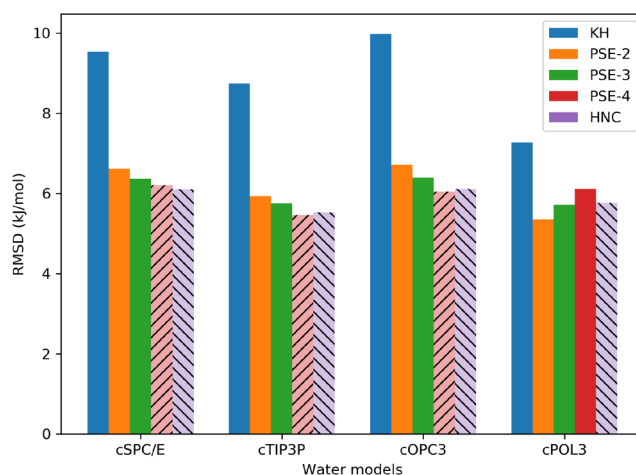


Figure 4. Comparison of RMSDs on the MNSOL[a] subset for different water models and closure relations. The paled and hatched bars represent the convergence issues with that pairing.

In summary, any of the studied water models can be used along with the PSE-2 or PSE-3 closure to get accurate solvation free energies. The PSE-3/cTIP3P and PSE-2/cPOL3 pairings are used in the next section to compare the method to other solvation models.

Level of Theory of QM Calculation. Besides MP2, HF, and several DFT functionals of different rungs (revPBE,⁸⁸ SCAN,⁸⁹ ω -B97X-V,⁹⁰ B2GPPLYP⁹¹) were also tested. Table 4 shows the corresponding statistics on the MNSOL[a] data set. HF performs worse than MP2 and DFT with an RMSD of 9.34 kJ/mol. The performance of the various DFT functionals is fairly similar with the RMSD ranging from 6.3 to 6.8 kJ/mol, which is worse than MP2 but still acceptable. Somewhat surprisingly, the more sophisticated double hybrid functional is also outperformed by MP2. This shows that the EC-RISM method is robust regarding the level of theory.

Comparison with the Original Implementation. We reproduced the results of the SAMPL6 paper on the MNSOL[a] subset⁴¹ within reasonably accuracy, knowing the differences between the RISM and CHELPG implementations. The main differences between our and the approach used in the SAMPL6 challenge (with atomic charges) are the following. Here, the QM method is MP2/def2-TZVPPD instead of MP2/6-311+G(d,p). We optimized the geometry in gas phase at the B3LYP/def2-TZVP level of theory, while Kast et al. used B3LYP/6-311+G(d,p) in gas phase and B3LYP/6-311+G(d,p)/IEFPCM in solution phase. They employed CHELPG charges fitted on the HF density and only calculated MP2 energy in the last step, while we take the density of the QM method applied throughout the whole calculation. Also, we used 1.75 Å as a radius for Br to determine CHELPG charges, while Kast et al. applied 1.3 Å. Additionally, they employed HF electrostatics to determine E_q^{QM} , while we used the corresponding MP2 electrostatics. For the 3D-RISM computations, we applied a water susceptibility function produced with the same closure that is otherwise utilized in the 3D-RISM calculation, while in the original version, the susceptibility function of mSPC/E water model⁷⁹ was employed with HNC closure. Also we apply PC+ correction, whereas Kast and co-workers utilized an UC-like correction, similar to eq 13, but a_0 was set to 0. For the 3D-RISM

Table 4. Error Statistics Using HF, MP2, and Various DFT Approaches as the QM Method in EC-RISM with PSE-3 Closure, cTIP3P Water Model, PC+ Correction, and def2-TZVPPD Basis Set on the MNSOL[a] Dataset

statistics	HF	MP2	revPBE	SCAN	ω -B97X-V	B2GPPLYP
MSD	-5.44	-1.56	0.13	-1.91	-2.60	-2.08
MAD	6.81	4.40	5.58	4.96	4.98	4.84
RMSD	9.34	5.75	6.82	6.33	6.60	6.31

Table 5. Error Statistics on the MNSOL[a] Dataset in kJ/mol along with c_V Correction Factors in kJ/mol/Å³ Using Different Settings to Compare Our and the Original SAMPL6 Implementation^a

ESP	water model	correction	6-311+G**				def2-TZVPPD			
			MSD	MAD	RMSD	c_V	MSD	MAD	RMSD	c_V
HF	mSPC/E	UC	-0.813	4.729	6.359	-0.42977	-0.522	4.328	5.754	-0.43269
HF	mSPC/E	PC+	-5.577	6.458	8.152	-0.46081	-4.846	5.802	7.385	-0.46081
HF	cSPC/E	UC	-0.801	4.789	6.430	-0.49790	-0.509	4.402	5.838	-0.50087
HF	cSPC/E	PC+	-2.890	5.267	6.809	-0.51181	-2.155	4.715	6.099	-0.51181
HF	cTIP3P	UC	-0.699	4.531	5.990	-0.47846	-0.410	4.242	5.512	-0.48141
HF	cTIP3P	PC+	-1.943	4.724	6.136	-0.48673	-1.212	4.339	5.578	-0.48673
MP2	mSPC/E	UC	-0.924	5.463	7.255	-0.41517	-0.618	4.871	6.474	-0.42076
MP2	mSPC/E	PC+	-7.974	8.472	10.468	-0.46081	-6.812	7.399	9.266	-0.46081
MP2	cSPC/E	UC	-0.883	5.529	7.366	-0.48210	-0.617	4.948	6.570	-0.48829
MP2	cSPC/E	PC+	-5.371	6.918	8.797	-0.51181	-4.172	5.935	7.594	-0.51181
MP2	cTIP3P	UC	-0.765	5.067	6.736	-0.46397	-0.500	4.639	6.063	-0.46997
MP2	cTIP3P	PC+	-4.209	6.061	7.680	-0.48673	-3.039	5.201	6.644	-0.48673

^aESP refers to the electrostatics used for the determination of CHELPG charges and E_q energy. Gas and solvent phase geometries are from the SAMPL6 work of Kast et al.,⁴¹ which have been corrected.⁹² 1D-RISM calculations for solvent susceptibility file were made with HNC closure, while 3D-RISM calculations with PSE-2 closure. The GAFF force field was utilized, and the radius of Br was set to 1.3 Å.

calculations, we applied the GAFF2 force field, instead of GAFF used by Kast et al.

Beyond the reproduction of the SAMPL6 results, we compared the effects of the different electrostatics (HF or MP2), basis sets (6-311+G** or def2-TZVPPD), and water model (cSPC/E, cTIP3P, or mSPC/E) on the MNSOL[a] data set using the setup of the Kast group, that is, geometries from the SAMPL6 paper, 1D-RISM with HNC, 3D-RISM with PSE-2, GAFF force field and Br radius set to 1.3 Å. The comparison can be seen in Table 5.

Obviously, the fitted UC correction performs better than the PC+ correction, as stated earlier. The effect of the basis set is also clear, the larger and more reliable def2-TZVPPD produces better error statistics than 6-311+G** in every case. The water model used by Kast et al. with UC correction and HF electrostatics, that is, the original EC-RISM setup with atomic charges for SAMPL6, provides similar results, RMSDs of 6.359 and 5.754 kJ/mol with 6-311+G** and def2-TZVPPD, respectively, while with the cSPC/E water model, RMSDs of 6.430 and 5.838 kJ/mol are obtained, which are slightly worse statistics than with the cTIP3P water model, RMSDs of 5.990 and 5.512 kJ/mol. However, with PC+ correction, we got more negative solvation free energies with the mSPC/E water model than with the others, leading to worse error statistics. A similar conclusion can be drawn comparing the statistics of MP2 electrostatics of the water models to each other.

Surprisingly, HF electrostatics performs better than MP2 electrostatics on the MNSOL[a] data set. MSD shows that MP2 electrostatics lowers the solvation free energies, which are already lower than the reference on average, and that leads to worse performance. Even though, it seems better to use HF electrostatics, we continue to employ our approach, that is, to use the same electrostatics for the CHELPG charges and the E_q energy as the QM level of theory because we think that it is

problematic to use the interaction energy of the HF density and the point charges to correct the MP2 energy instead of the interaction energy of MP2 density and the point charges. However, the usage of HF electrostatics can save around 30–50% computational time in the intermediate cycles by not calculating the MP2 energy and density, only at the last cycle.

In summary, any of the water models can be used with the UC correction, and the results can be improved with larger basis sets, however, cTIP3P performs better with PC+. Also, based on theoretical considerations, we prefer MP2 electrostatics over HF electrostatics, even though the latter gives better error statistics in every studied setup.

Comparison with Other Solvation Models. The best two settings, PSE-3/cTIP3P and PSE-2/cPOL3, were tested on both the full MNSOL and the FreeSolv databases with the relevant conformers taken into account, while other settings were kept as determined earlier: MP2/def2-TZVPPD, PC+ correction, 15 Å buffer, 0.3 Å grid space, CHELPG charges. We are aware of the fact that the good performance of a model on these data sets does not always lead to successful prediction of solvation parameters in general, as the SAMPL challenges showed.^{93–96} However, without studying other solvents and ionic species, which will be addressed in future works, we cannot test the method on the SAMPL challenges and have to rely on the MNSOL and FreeSolv data sets.

For the MNSOL[a] database, we got slightly better RMSDs considering the conformers (5.70 and 5.22 kJ/mol) than originally (5.75 and 5.35 kJ/mol). The full MNSOL database gave higher RMSDs with 6.32 and 6.08 kJ/mol. However, for the FreeSolv database PSE-3/cTIP3P performed slightly better: an RMSD of 6.44 kJ/mol was obtained with this combination of settings, while PSE-2/cPOL3 yielded 6.63 kJ/mol.

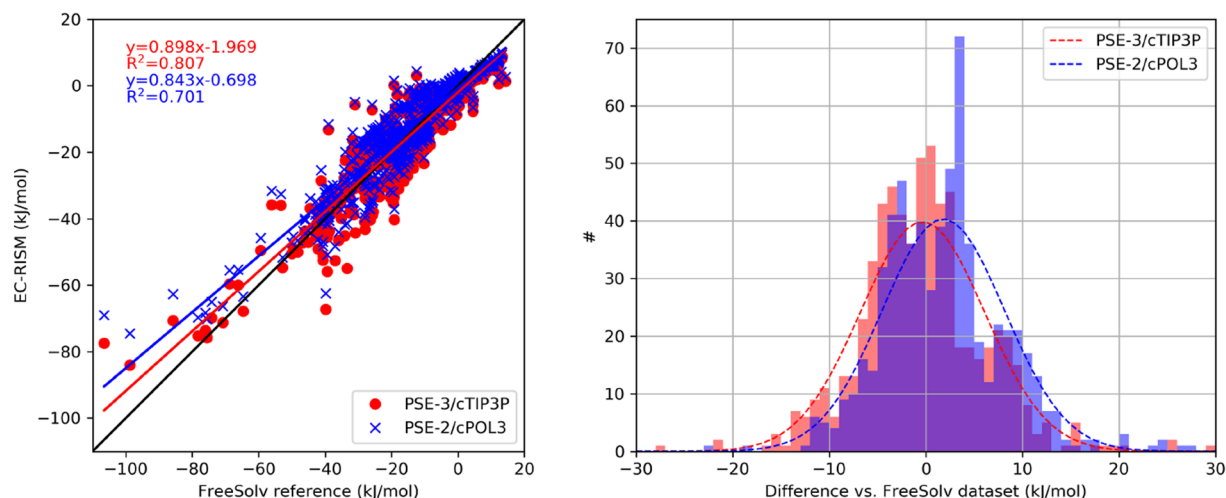


Figure 5. Calculated vs reference experimental solvation free energies (left) and histogram of errors (right) on the FreeSolv data set for PSE-3/cTIP3P and PSE-2/cPOL3 EC-RISM settings.

Table 6. Comparison of RMSDs of Several Solvation Models on the MNSOL[a], MNSOL, and FreeSolv Satasets^a

method	FreeSolv	MNSOL[a]	MNSOL
EC-RISM, MP2/def2-TZVPPD/ PSE-3/cTIP3P/CHELPG/PC+, this work	6.44	5.70	6.32
EC-RISM, MP2/def2-TZVPPD/ PSE-2/cPOL3/CHELPG/PC+, this work	6.63	5.22	6.08
EC-RISM, MP2/6-311+G(d,p)/ PSE-2/mSPC/E/ φ_{opt} /UC ⁴¹	—	6.19	6.53
EC-RISM, MP2/6-311+G(d,p)/ PSE-2/mSPC/E/CHELPG/UC ⁴¹	—	6.59	7.40
EC-RISM, MP2/cc-pVTZ/ PSE-2/mSPC/E/ φ_{opt} /UC ⁴¹	—	4.73	5.44
3D-RISM, PSE-4/cTIP3P/AM1-BCC, this work	7.12	6.33	8.18
3D-RISM, PSE-3/cPOL3/AM1-BCC, this work	7.39	5.92	7.50
MD ^{4,45}	6.45	—	—
DCOSMO-RS ϵ ¹¹	—	5.93	—
DCOSMO-RS ∞ ¹¹	—	6.80	—
COSMO ¹¹	—	10.42	—
COSMO-RS ^{11,97}	—	3.48	—
SMD, IEF-PCM/G03/M05-2x ¹²	—	3.60	—
SMD, M05-2X/6-31G(d), this work	5.90	4.20	6.85
IEF-PCM, G03d(UAHF)/HF ¹²	—	7.32	—
PCM, M05-2X/6-31G(d), this work	9.00	7.95	9.53
C-PCM, G06d/B3LYP ¹²	—	7.74	—

^aValues are in kJ/mol.

Figure 5 compares the calculated and reference solvation free energies and presents the histogram of the differences. We can see that larger deviations from the reference values occur for the more negative solvation free energies, while the distribution is close to the desired Gaussian distribution for PSE-3/cTIP3P with the FreeSolv data set. However, for PSE-2/cPOL3 the distribution is skewed, the peak of the histogram is at 3–4 kJ/mol.

The most problematic species can be seen in Table S4. We cannot determine a clear reason why the EC-RISM results deviate from the reference values for the outliers. Some of them, D-xylose, D-glucose, mannitol, and glycerol, have numerous hydroxyl groups, which can be challenging to taken into account properly. Also druglike molecules (nitralin, trichlorfon, flufenamicacid) have several different functional

groups. It is surprising that such simple species as 2-nitrophenol and 2-iodophenol can cause difficulties for this method and result in a 20 kJ/mol error. Problems can also arise due to the improper atom type assignment during the force field application. Also, it is possible that the accuracy and reliability of some data is questionable.

We also compared the performance of our EC-RISM method to other solvation models. The RMSDs of the methods on the MNSOL, MNSOL[a] subset, and FreeSolv databases are collected in Table 6, while additional statistics with other methods are listed in Table S5.

On the MNSOL[a] data set, the COSMO-RS and the SMD solvation models perform the best with RMSDs of under 4 kJ/mol, but it must be noted that they were optimized for this data set. The EC-RISM implementation of Kast et al.,⁴¹ where,

instead of CHELPG charges, the electrostatic potential is used directly in the RISM calculations, is capable of providing an RMSD of 4.73 kJ/mol. This shows that our implementation has still room for improvements.

For a better comparison for SMD, PCM, and 3D-RISM, we performed calculations for the whole FreeSolv and MNSOL databases. For 3D-RISM, we used similar settings as for EC-RISM, but AM1-BCC charges were employed. For SMD and PCM, we used the M05-2X/6-31G(d) level of theory. The results are varying for SMD: though it performs slightly worse than the reference on MNSOL[a], it is better than our EC-RISM on the FreeSolv database (RMSDs of 5.90 vs 6.44 kJ/mol) but worse on the whole MNSOL test set (6.85 vs 6.32 kJ/mol).

Comparing our and Kast's EC-RISM, the former performs better on the whole MNSOL data set, except when the cc-pVTZ basis set and the direct electrostatic potential are utilized. The reason for that is that they use UC which is fitted on the MNSOL[a] data set, while our implementation is free from such empirical parameters. Another difference is that they use HF during the iterations, and MP2 is only employed for the converged state, which can speed up the calculation but can have the opposite effect on the accuracy.

If we look at the pure 3D-RISM, we can see that PSE-3/cPOL3 is the best performing approach on the MNSOL[a] and MNSOL data sets (RMSDs of 5.92 and 7.50 kJ/mol), while PSE-4/cTIP3P is better on the FreeSolv database with an RMSD of 7.12 kJ/mol. That is, 3D-RISM is slightly worse than our EC-RISM but still provides comparable solvation free energies. The COSMO and PCM models are not reliable for the calculation of solvation free energies. The MD results provided by Mobley for their FreeSolv data set show 6.45 kJ/mol as RMSD, which is similar to our 6.44 kJ/mol for our EC-RISM result with PSE-3/cTIP3P.

On the basis of these results, it is hard to say which is the best solvation model, but besides COSMO-RS, our method is definitely comparable with the original EC-RISM of Kast, MD, and SMD solvation models. Even though EC-RISM is more expensive than implicit solvent models due to the fact that it requires 2–8 more QM calculations to converge, RISM can provide additional information about solvent distribution. Further optimization and testing is needed for ionic solutes and nonwater solvents, which will be addressed in future studies.

CONCLUSIONS

In this work, we implemented the EC-RISM method of Kloss, Heil, and Kast³⁷ with atomic charges and optimized various settings of it. We have found that the buffer should be at least 10 Å, but we recommend 15 Å. For grid space, we chose 0.3 Å, but it should be less than 0.5 Å. The solvent distribution of RISM is represented as point charges on a grid, and the small point charges can be discarded to reduce the overhead of the QM calculation. For the threshold for neglecting the point charges, we suggest 10^{-6} a.u., but 10^{-5} a.u. is still acceptable. Regarding the basis sets, we found that for EC-RISM calculations, the usage of diffuse functions is recommended, and def2-TZVPPD was utilized throughout the work. As for the charge models, the ones based on the electrostatic potential, like CHELPG and MK, provide acceptable results, and we chose CHELPG as default. Out of the several corrections for 3D-RISM, only PC+ and UC provide meaningful results regarding solvation free energy, and we

recommend to use PC+ because it does not require a training set to determine the fitting parameters for other settings. The level of the QM calculation can be almost anything, MP2 is slightly better than other DFT alternatives. The choice of the water model has only a small effect, but cTIP3P and cPOL3 perform better than cSPC/E and cOCP3. The closure relation for 3D-RISM should be PSE-2 or PSE-3 with PSE-3/cTIP3P and PSE-2/cPOL3 being the best pairings. However, PSE-3/cTIP3P has been chosen as default because it performs better on the largest FreeSolv data set, and a cTIP3P water model file is already available in the recent AMBERTOOLS versions.

On the MNSOL[a], MNSOL, and FreeSolv databases, we got 5.70, 6.32, and 6.44 kJ/mol RMSDs for the optimized settings with PSE-3/cTIP3P, while PSE-2/cPOL3 produced 5.22, 6.08, and 6.63 kJ/mol, respectively. Compared to other solvation models, COSMO-RS has the best performance on the MNSOL[a] subset (3.48 kJ/mol),^{11,97} but it was optimized on that data set, and we have no data for the other data sets. The EC-RISM of Kast et al. using the electrostatic potential directly and the UC correction also performs better on the MNSOL database than our implementation (RMSD of 5.44 kJ/mol), but with different basis set and/or with CHELPG charges, slightly worse RMSDs, 6.53 and 7.40 kJ/mol, can be obtained. SMD, which is also fitted on the MNSOL[a] subset, provides better statistics on the MNSOL[a] and FreeSolv data sets but worse on MNSOL. Pure 3D-RISM produces acceptable solvation free energies, but the RMSDs are obviously worse than the EC-RISM ones. The performance of MD is comparable to our implemented method. In summary, only solvent models containing fitting parameters, COSMO-RS and EC-RISM with universal correction, perform better than our EC-RISM implementation.

We note that our approach could be improved with the direct usage of the electrostatic potential in the RISM calculations⁴¹ and also with better force fields, especially with polarizable ones.³⁹ The methodology can also be extended to ions, which requires an explicit charge correction term, and nonwater solvents as suggested by the EC-RISM results in the SAMPL challenges,⁴¹ and this is planned for future work.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpca.1c07904>.

Solvation free energies and statistics on the MNSOL[a], MNSOL, and FreeSolv databases; geometries for all the systems discussed in this paper. Also raw data is provided in csv format including the names of the species in the MNSOL and FreeSolv databases, reference values, uncorrected 3D-RISM solvation free energies, PMV, corrections to 3D-RISM, corrected 3D-RISM results, gas-phase electronic energies, electronic energies in solution, and the EC-RISM solvation free energies (ZIP)

AUTHOR INFORMATION

Corresponding Author

Ádám Ganyecz – Department of Physical Chemistry and Materials Science, Budapest University of Technology and Economics, Budapest H-1521, Hungary; orcid.org/0000-0002-8828-0283; Email: ganyecz.adam@vbk.bme.hu

Author

Mihály Kállay – Department of Physical Chemistry and Materials Science, Budapest University of Technology and Economics, Budapest H-1521, Hungary; orcid.org/0000-0003-1080-6625

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jpca.1c07904>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The research within project No. VEKOP-2.3.2-16-2017-00013 was supported by the European Union and the State of Hungary, cofinanced by the European Regional Development Fund. The authors are also grateful for the financial support from the National Research, Development, and Innovation Office (NKFIH, Grant No. KKP126451). The computing time granted on the Hungarian HPC Infrastructure at NIIIF Institute, Hungary, is gratefully acknowledged. We would like to thank Stefan Kast and Nicolas Tielker for their help to reproduce their results and to compare the different EC-RISM implementations.

REFERENCES

- (1) Ratkova, E. L.; Palmer, D. S.; Fedorov, M. V. Solvation Thermodynamics of Organic Molecules by the Molecular Integral Equation Theory: Approaching Chemical Accuracy. *Chem. Rev.* **2015**, *115*, 6312–6356.
- (2) Van De Waterbeemd, H.; Gifford, E. ADMET in silico Modelling: Towards Prediction Paradise? *Nat. Rev. Drug Discovery* **2003**, *2*, 192–204.
- (3) Skyner, R.; McDonagh, J.; Groom, C.; van Mourik, T.; Mitchell, J. A Review of Methods for the Calculation of Solution Free Energies and the Modelling of Systems in Solution. *Phys. Chem. Chem. Phys.* **2015**, *17*, 6174–6191.
- (4) Mobley, D. L.; Guthrie, J. P. FreeSolv: A Database of Experimental and Calculated Hydration Free Energies, with Input Files. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 711–720.
- (5) Miertuš, S.; Scrocco, E.; Tomasi, J. Electrostatic Interaction of a Solute with a Continuum. A Direct Utilization of ab initio Molecular Potentials for the Prediction of Solvent Effects. *Chem. Phys.* **1981**, *55*, 117–129.
- (6) Cancès, E.; Mennucci, B.; Tomasi, J. A New Integral Equation Formalism for the Polarizable Continuum Model: Theoretical Background and Applications to Isotropic and Anisotropic Dielectrics. *J. Chem. Phys.* **1997**, *107*, 3032–3041.
- (7) Mennucci, B.; Cammi, R.; Tomasi, J. Excited States and Solvatochromic Shifts Within a Nonequilibrium Solvation Approach: A New Formulation of the Integral Equation Formalism Method at the Self-Consistent Field, Configuration Interaction, and Multi-configuration Self-Consistent Field Level. *J. Chem. Phys.* **1998**, *109*, 2798–2807.
- (8) Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998**, *102*, 1995–2001.
- (9) Klamt, A.; Schüürmann, G. COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and its Gradient. *J. Chem. Soc., Perkin Trans.* **1993**, *2*, 799–805.
- (10) Klamt, A. *COSMO-RS: From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*; Elsevier: Amsterdam, 2005.
- (11) Klamt, A.; Diedenhofen, M. Calculation of Solvation Free Energies with DCOSMO-RS. *J. Phys. Chem. A* **2015**, *119*, 5439–5445.
- (12) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.
- (13) Mennucci, B.; Cammi, R. *Continuum Solvation Models in Chemical Physics: From Theory to Applications*; John Wiley & Sons: Hoboken, NJ, 2008.
- (14) Chandler, D.; Andersen, H. C. Optimized Cluster Expansions for Classical Fluids. II. Theory of Molecular Liquids. *J. Chem. Phys.* **1972**, *57*, 1930–1937.
- (15) Beglov, D.; Roux, B. Solvation of Complex Molecules in a Polar Liquid: an Integral Equation Theory. *J. Chem. Phys.* **1996**, *104*, 8678–8689.
- (16) Beglov, D.; Roux, B. An Integral Equation to Describe the Solvation of Polar Molecules in Liquid Water. *J. Phys. Chem. B* **1997**, *101*, 7821–7826.
- (17) Ramirez, R.; Gebauer, R.; Mareschal, M.; Borgis, D. Density Functional Theory of Solvation in a Polar Solvent: Extracting the Functional from Homogeneous Solvent Simulations. *Phys. Rev. E* **2002**, *66*, 031206.
- (18) Ramirez, R.; Borgis, D. Density Functional Theory of Solvation and its Relation to Implicit Solvent Models. *J. Phys. Chem. B* **2005**, *109*, 6754–6763.
- (19) Ramirez, R.; Mareschal, M.; Borgis, D. Direct Correlation Functions and the Density Functional Theory of Polar Solvents. *Chem. Phys.* **2005**, *319*, 261–272.
- (20) Gendre, L.; Ramirez, R.; Borgis, D. Classical Density Functional Theory of Solvation in Molecular Solvents: Angular Grid Implementation. *Chem. Phys. Lett.* **2009**, *474*, 366–370.
- (21) Zhao, S.; Ramirez, R.; Vuilleumier, R.; Borgis, D. Molecular Density Functional Theory of Solvation: From Polar Solvents to Water. *J. Chem. Phys.* **2011**, *134*, 194102.
- (22) Borgis, D.; Gendre, L.; Ramirez, R. Molecular Density Functional Theory: Application to Solvation and Electron-Transfer Thermodynamics in Polar Solvents. *J. Phys. Chem. B* **2012**, *116*, 2504–2512.
- (23) Beglov, D.; Roux, B. Numerical Solution of the Hypernetted Chain Equation for a Solute of Arbitrary Geometry in Three Dimensions. *J. Chem. Phys.* **1995**, *103*, 360–364.
- (24) Hirata, F. *Molecular Theory of Solvation*; Springer Science & Business Media: Berlin, Heidelberg, 2003.
- (25) 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.
- (26) 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.
- (27) Kast, S. M.; Kloss, T. Closed-Form Expressions of the Chemical Potential for Integral Equation Closures with Certain Bridge Functions. *J. Chem. Phys.* **2008**, *129*, 236101.
- (28) Palmer, D. S.; Frolov, A. I.; Ratkova, E. L.; Fedorov, M. V. Towards a Universal Method for Calculating Hydration Free Energies: A 3D Reference Interaction Site Model with Partial Molar Volume Correction. *J. Phys.: Condens. Matter* **2010**, *22*, 492101.
- (29) Sergiievskiy, V. P.; Jeanmairet, G.; Levesque, M.; Borgis, D. Fast Computation of Solvation Free Energies with Molecular Density Functional Theory: Thermodynamic-Ensemble Partial Molar Volume corrections. *J. Phys. Chem. Lett.* **2014**, *5*, 1935–1942.
- (30) Misin, M.; Fedorov, M. V.; Palmer, D. S. Communication: Accurate Hydration Free Energies at a Wide Range of Temperatures from 3D-RISM. *J. Chem. Phys.* **2015**, *142*, 091105.
- (31) Sergiievskiy, V.; Jeanmairet, G.; Levesque, M.; Borgis, D. Solvation Free-Energy Pressure Corrections in the Three Dimensional Reference Interaction Site Model. *J. Chem. Phys.* **2015**, *143*, 184116.
- (32) Ten-no, S.; Hirata, F.; Kato, S. A Hybrid Approach for the Solvent Effect on the Electronic Structure of a Solute Based on the RISM and Hartree-Fock Equations. *Chem. Phys. Lett.* **1993**, *214*, 391–396.

- (33) Yokogawa, D.; Sato, H.; Sakaki, S. New Generation of the Reference Interaction Site Model Self-Consistent Field Method: Introduction of Spatial Electron Density Distribution to the Solvation Theory. *J. Chem. Phys.* **2007**, *126*, 244504.
- (34) Sato, H.; Kovalenko, A.; Hirata, F. Self-Consistent Field, ab initio Molecular Orbital and Three-Dimensional Reference Interaction Site Model Study for Solvation Effect on Carbon Monoxide in Aqueous Solution. *J. Chem. Phys.* **2000**, *112*, 9463–9468.
- (35) Gusarov, S.; Ziegler, T.; Kovalenko, A. Self-Consistent Combination of the Three-Dimensional RISM Theory of Molecular Solvation with Analytical Gradients and the Amsterdam Density Functional Package. *J. Phys. Chem. A* **2006**, *110*, 6083–6090.
- (36) 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.
- (37) Kloss, T.; Heil, J.; Kast, S. M. Quantum Chemistry in Solution by Combining 3D Integral Equation Theory with a Cluster Embedding Approach. *J. Phys. Chem. B* **2008**, *112*, 4337–4343.
- (38) Case, D.; Aktulga, H.; Belfon, K.; Ben-Shalom, I.; Brozell, S.; Cerutti, D.; Cheatham, T., III; Cruzeiro, V.; Darden, T.; Duke, R. et al. *AMBER*, 2021.
- (39) Hoffgaard, F.; Heil, J.; Kast, S. M. Three-Dimensional RISM Integral Equation Theory for Polarizable Solute Models. *J. Chem. Theor. Comp.* **2013**, *9*, 4718–4726.
- (40) Frach, R.; Kast, S. M. Solvation Effects on Chemical Shifts by Embedded Cluster Integral Equation Theory. *J. Phys. Chem. A* **2014**, *118*, 11620–11628.
- (41) Tielker, N.; Eberlein, L.; Güssregen, S.; Kast, S. M. The SAMPL6 Challenge on Predicting Aqueous pKa Values from EC-RISM Theory. *J. Comput.-Aided Mol. Des.* **2018**, *32*, 1151–1163.
- (42) Tielker, N.; Güssregen, S.; Kast, S. M. SAMPL7 Physical Property Prediction from EC-RISM Theory. *J. Comput.-Aided Mol. Des.* **2021**, *35*, 933–941.
- (43) Tielker, N.; Eberlein, L.; Hessler, G.; Schmidt, K. F.; Güssregen, S.; Kast, S. M. Quantum-Mechanical Property Prediction of Solvated Drug Molecules: What Have We Learned From a Decade of SAMPL Blind Prediction Challenges? *J. Comput.-Aided Mol. Des.* **2021**, *35*, 453–472.
- (44) Marenich, A. V.; Kelly, C. P.; Thompson, J. D.; Hawkins, G. D.; Chambers, C. C.; Giesen, D. J.; Winget, P.; Cramer, C. J.; Truhlar, D. G. *Minnesota Solvation Database (MNSOL)*, version 2012.
- (45) Duarte Ramos Matos, G.; Kyu, D. Y.; Loeffler, H. H.; Chodera, J. D.; Shirts, M. R.; Mobley, D. L. Approaches for Calculating Solvation Free Energies and Enthalpies Demonstrated with an Update of the FreeSolv Database. *J. Chem. Eng. Data* **2017**, *62*, 1559–1569.
- (46) Monson, P.; Morriss, G. Recent Progress in the Statistical Mechanics of Interaction Site Fluids. *Adv. Chem. Phys.* **2007**, *77*, 451–550.
- (47) Barrat, J.-L.; Hansen, J.-P. *Basic Concepts for Simple and Complex Liquids*; Cambridge University Press: Cambridge, 2003.
- (48) Ornstein, L. S.; Zernike, F. Integral Equation in Liquid State Theory. *Proc. Acad. Sci. Amsterdam* **1914**, *17*, 793.
- (49) Hansen, J.-P.; McDonald, I. R. *Theory of Simple Liquids*; Elsevier: Amsterdam, 1990.
- (50) Hirata, F.; Rossky, P. J. An Extended RISM Equation for Molecular Polar Fluids. *Chem. Phys. Lett.* **1981**, *83*, 329–334.
- (51) Cummings, P. T.; Stell, G. Exact Asymptotic Form of the Site-Site Direct Correlation Function for Rigid Polar Molecules. *Mol. Phys.* **1981**, *44*, 529–531.
- (52) Perkyns, J.; Pettitt, B. M. A Site-Site Theory for Finite Concentration Saline Solutions. *J. Chem. Phys.* **1992**, *97*, 7656–7666.
- (53) Kovalenko, A.; Hirata, F. Three-dimensional Density Profiles of Water in Contact with a Solute of Arbitrary Shape: a RISM Approach. *Chem. Phys. Lett.* **1998**, *290*, 237–244.
- (54) Singer, S. J.; Chandler, D. Free Energy Functions in the Extended RISM Approximation. *Mol. Phys.* **1985**, *55*, 621–625.
- (55) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*; Elsevier: Amsterdam, 2001.
- (56) Chandler, D.; Singh, Y.; Richardson, D. M. Excess Electrons in Simple Fluids. I. General Equilibrium Theory for Classical Hard Sphere Solvents. *J. Chem. Phys.* **1984**, *81*, 1975–1982.
- (57) 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.
- (58) 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.
- (59) Kast, S. M.; Heil, J.; Güssregen, S.; Schmidt, K. F. Prediction of Tautomer Ratios by Embedded-Cluster Integral Equation Theory. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 343–353.
- (60) Kállay, M.; Nagy, P. R.; Mester, D.; Rolik, Z.; Samu, G.; Csontos, J.; Csóka, J.; Szabó, P. B.; Gyevi-Nagy, L.; Hégyel, B.; et al. The MRCC program system: Accurate quantum chemistry from water to proteins. *J. Chem. Phys.* **2020**, *152*, 074107.
- (61) Misin, M.; Fedorov, M. V.; Palmer, D. S. Hydration Free Energies of Molecular Ions from Theory and Simulation. *J. Phys. Chem. B* **2016**, *120*, 975–983.
- (62) Tielker, N.; Tomazic, D.; Heil, J.; Kloss, T.; Ehrhart, S.; Güssregen, S.; Schmidt, K. F.; Kast, S. M. The SAMPL5 Challenge for Embedded-Cluster Integral Equation Theory: Solvation Free Energies, Aqueous pKa, and Cyclohexane-Water log D. *J. Comput.-Aided Mol. Des.* **2016**, *30*, 1035–1044.
- (63) Pongratz, T.; Kibies, P.; Eberlein, L.; Tielker, N.; Holzl, C.; Imoto, S.; Beck Erlach, M.; Kurrmann, S.; Schummel, P. H.; Hofmann, M.; Reiser, O.; Winter, R.; Kremer, W.; Kalbitzer, H. R.; Marx, D.; Horinek, D.; Kast, S. M.; et al. Pressure-Dependent Electronic Structure Calculations Using Integral Equation-Based Solvation Models. *Biophys. Chem.* **2020**, *257*, 106258.
- (64) Hölzl, C.; Kibies, P.; Imoto, S.; Frach, R.; Suladze, S.; Winter, R.; Marx, D.; Horinek, D.; Kast, S. M. Design Principles for High-Pressure Force Fields: Aqueous TMAO Solutions from Ambient to Kilobar Pressures. *J. Chem. Phys.* **2016**, *144*, 144104.
- (65) Frach, R.; Kibies, P.; Böttcher, S.; Pongratz, T.; Strohfeldt, S.; Kurrmann, S.; Koehler, J.; Hofmann, M.; Kremer, W.; Kalbitzer, H. R.; et al. The Chemical Shift Baseline for High-Pressure NMR Spectra of Proteins. *Angew. Chem., Int. Ed.* **2016**, *55*, 8757–8760.
- (66) Imoto, S.; Kibies, P.; Rosin, C.; Winter, R.; Kast, S. M.; Marx, D. Toward Extreme Biophysics: Deciphering the Infrared Response of Biomolecular Solutions at High Pressures. *Angew. Chem., Int. Ed.* **2016**, *55*, 9534–9538.
- (67) Hoelzl, C.; Kibies, P.; Imoto, S.; Noetzel, J.; Knierbein, M.; Salmen, P.; Paulus, M.; Nase, J.; Held, C.; Sadowski, G.; et al. Structure and Thermodynamics of Aqueous Urea Solutions from Ambient to Kilobar Pressures: From Thermodynamic Modeling, Experiments, and First Principles Simulations to an Accurate Force Field Description. *Biophys. Chem.* **2019**, *254*, 106260.
- (68) O’Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminformatics* **2011**, *3*, 1–14.
- (69) Dunning, T. H., Jr. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007.
- (70) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. Electron Affinities of the First-Row Atoms Revisited. Systematic Basis Sets and Wave Functions. *J. Chem. Phys.* **1992**, *96*, 6796.
- (71) Woon, D. E.; Dunning, T. H., Jr. Gaussian Basis Sets for Use in Correlated Molecular Calculations. III. The Atoms Aluminum through Argon. *J. Chem. Phys.* **1993**, *98*, 1358.
- (72) Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
- (73) Berendsen, H.; Grigera, J.; Straatsma, T. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

- (74) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (75) Izadi, S.; Onufriev, A. V. Accuracy Limit of Rigid 3-point Water Models. *J. Chem. Phys.* **2016**, *145*, 074501.
- (76) Caldwell, J. W.; Kollman, P. A. Structure and Properties of Neat Liquids Using Nonadditive Molecular Dynamics: Water, Methanol, and N-methylacetamide. *J. Phys. Chem.* **1995**, *99*, 6208–6219.
- (77) Pettitt, B. M.; Rossky, P. J. Integral Equation Predictions of Liquid State Structure for Waterlike Intermolecular Potentials. *J. Chem. Phys.* **1982**, *77*, 1451–1457.
- (78) Hirata, F.; Levy, R. M. A New RISM Integral Equation for Solvated Polymers. *Chem. Phys. Lett.* **1987**, *136*, 267–273.
- (79) Maw, S.; Sato, H.; Ten-no, S.; Hirata, F. Ab initio study of water: self-consistent determination of electronic structure and liquid state properties. *Chemical physics letters* **1997**, *276*, 20–25.
- (80) 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. Theor. Comp.* **2010**, *6*, 607–624.
- (81) Dong, X.; Yuan, X.; Song, Z.; Wang, Q. The Development of an Amber-compatible Organosilane Force Field for Drug-Like Small Molecules. *Phys. Chem. Chem. Phys.* **2021**, *23*, 12582.
- (82) Breneman, C. M.; Wiberg, K. B. Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials. The Need for High Sampling Density in Formamide Conformational Analysis. *J. Comput. Chem.* **1990**, *11*, 361–373.
- (83) Singh, U. C.; Kollman, P. A. An Approach to Computing Electrostatic Charges for Molecules. *J. Comput. Chem.* **1984**, *5*, 129–145.
- (84) Knizia, G. Intrinsic Atomic Orbitals: An Unbiased Bridge between Quantum Theory and Chemical Concepts. *J. Chem. Theor. Comp.* **2013**, *9*, 4834.
- (85) Mulliken, R. S. Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I. *J. Chem. Phys.* **1955**, *23*, 1833.
- (86) Löwdin, P.-O. On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals. *J. Chem. Phys.* **1950**, *18*, 365–375.
- (87) Roy, D.; Kovalenko, A. Performance of 3D-RISM-KH in Predicting Hydration Free Energy: Effect of Solute Parameters. *J. Phys. Chem. A* **2019**, *123*, 4087–4093.
- (88) Zhang, Y.; Yang, W. Comment on “Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1998**, *80*, 890–890.
- (89) Sun, J.; Ruzsinszky, A.; Perdew, J. P. Strongly Constrained and Appropriately Normed Semilocal Density Functional. *Phys. Rev. Lett.* **2015**, *115*, 036402.
- (90) Mardirossian, N.; Head-Gordon, M. ω B97X-V: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy. *Phys. Chem. Chem. Phys.* **2014**, *16*, 9904.
- (91) Karton, A.; Tarnopolsky, A.; Lamère, J.-F.; Schatz, G. C.; Martin, J. M. L. Highly Accurate First-Principles Benchmark Data Sets for the Parametrization and Validation of Density Functional and Other Approximate Methods. Derivation of a Robust, Generally Applicable, Double-Hybrid Functional for Thermochemistry and Thermochemical Kinetics. *J. Phys. Chem. A* **2008**, *112*, 12868.
- (92) Kast, S. M.; Tielker, N. personal communication, 2022.
- (93) Mobley, D. L.; Wymer, K. L.; Lim, N. M.; Guthrie, J. P. Blind Prediction of Solvation Free Energies from the SAMPL4 Challenge. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 135–150.
- (94) Bannan, C. C.; Burley, K. H.; Chiu, M.; Shirts, M. R.; Gilson, M. K.; Mobley, D. L. Blind Prediction of Cyclohexane-Water Distribution Coefficients from the SAMPL5 Challenge. *J. Comput.-Aided Mol. Des.* **2016**, *30*, 927–944.
- (95) Işık, M.; Bergazin, T. D.; Fox, T.; Rizzi, A.; Chodera, J. D.; Mobley, D. L. Assessing the Accuracy of Octanol-Water Partition Coefficient Predictions in the SAMPL6 Part II log P Challenge. *J. Comput.-Aided Mol. Des.* **2020**, *34*, 335–370.
- (96) Bergazin, T. D.; Tielker, N.; Zhang, Y.; Mao, J.; Gunner, M. R.; Francisco, K.; Ballatore, C.; Kast, S. M.; Mobley, D. L. Evaluation of log P, pKa, and log D predictions from the SAMPL7 blind challenge. *J. Comput.-Aided Mol. Des.* **2021**, *35*, 771–802.
- (97) Klamt, A. Conductor-Like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.* **1995**, *99*, 2224–2235.