

Accurate hydration free energies at a wide range of temperatures from 3D-RISM

Maksim Misin,¹ Maxim V. Fedorov,^{1, a)} and David S. Palmer^{2, b)}

¹⁾*Department of Physics, SUPA, University of Strathclyde, 107 Rottenrow, Glasgow, G4 0NG, UK*

²⁾*Department of Pure and Applied Chemistry, University of Strathclyde, 16 Richmond Street, Glasgow, G1 1XQ, UK*

(Dated: 23 February 2015)

We present a new model for computing hydration free energies by 3D-RISM that uses an appropriate initial state of the system (as suggested by Sergiievskiy et al.). The new adjustment to 3D-RISM theory significantly improves hydration free energy predictions for various classes of organic molecules at both ambient and non-ambient temperatures. An extensive benchmarking against experimental data shows that the accuracy of the model is comparable to (much more computationally expensive) molecular dynamics simulations. The calculations can be readily performed with a standard 3D-RISM algorithm. In our work we used an open source package AmberTools; a script to automate the whole procedure is available on the web (<https://github.com/MTS-Strathclyde/ISc>).

I. INTRODUCTION

Hydration free energy, ΔG_{hyd} , is one of the most important properties in solution chemistry. It provides information about the partitioning of a solute between gas and solution phases and is used in calculating solubility¹⁻³, acid-base dissociation constant⁴, octanol-water partition coefficient⁵, and protein-ligand binding free energy⁶, amongst other properties. The hydration free energies of non-ionic organic solutes can be computed using molecular dynamics (MD) simulations with an accuracy approaching that of experiments.⁷ However, for many purposes (in particular for high-throughput screening over large digital libraries of compounds) the MD approach is not ideal as it requires large computational resources⁷⁻⁹. Additionally, the large number of existing MD methods and the variety of input parameters both for the model itself (molecular geometry, force-field, number of molecules in the simulation box, etc) and for the simulation protocol (time step, MD integrator, thermostat, barostat, equilibration time etc) together with the need to use a parallel computer architecture makes it difficult for non-specialists to perform these calculations.

The 3D Reference Interaction Site Model (3D-RISM) is a molecular theory closely related to classical DFT¹⁰⁻¹³. It describes the local solvent density around a solute using 3D spatial solute-solvent distribution functions, which are obtained via iterative solution of a system of integral equations. Using 3D-RISM one can compute a range of thermodynamic properties as well as structural information about the solvent in a relatively short time (on the order of a few minutes)¹⁴⁻¹⁷. In principle it can be applied to arbitrary solvent compositions and temperatures^{18-20,72-74}. However, despite recent advances, such as the introduction of semi-empirical free

energy functionals,^{15,21-24} the accuracy of theory-based 3D-RISM methods (i.e. those without empirical corrections) has remained low²⁵.

In this paper we build upon recent theoretical work by Sergiievskiy *et al.*²⁵ and propose a new correction for the 3D-RISM hypernetted-chain (3D-RISM/HNC) theory. To assess the strengths and limitations of the new model, we compare its predictions to experimental hydration free energy data for small organic molecules at both ambient and non-ambient temperatures. We show that the new term not only improves predictions at 298 K, but also makes the error practically temperature-independent for the range of temperatures between 278 K and 378 K that were studied here.

II. THEORY

An important part of 3D-RISM theory is the so called closure expression. It provides a connection between intramolecular potential $u(\mathbf{r})$ and distribution functions:

$$h(\mathbf{r}) + 1 = \exp(-\beta u(\mathbf{r}) + h(\mathbf{r}) - c(\mathbf{r}) + B(\mathbf{r})) \quad (1)$$

here $h(\mathbf{r})$ is the total correlation function, $c(\mathbf{r})$ is the direct correlation function, and $B(\mathbf{r})$ is a bridge function, defined using diagrammatic expansion^{14,26-29}. In practice however, one often substitutes $B(\mathbf{r})$ with various approximations.

A common choice is to set $B(\mathbf{r}) = 0$ in the so called hypernetted-chain approximation (HNC). However, this approach suffers from poor computational convergence^{30,31}. This problem can be reduced by the use of a partial series expansion of order n (PSE- n) of the HNC closure³⁰:

$$h(\mathbf{r}) = \begin{cases} \sum_{i=0}^n \Xi^i(\mathbf{r})/i! - 1 & \text{if } \Xi(\mathbf{r}) > 0 \\ \exp(\Xi(\mathbf{r})) - 1 & \text{if } \Xi(\mathbf{r}) \leq 0 \end{cases} \quad (2)$$

^{a)}Electronic mail: maxim.fedorov@strath.ac.uk

^{b)}Electronic mail: david.palmer@strath.ac.uk

where $\Xi(\mathbf{r}) = -\beta u(\mathbf{r}) + h(\mathbf{r}) - c(\mathbf{r})$. Setting $n = 1$ results in the partial linearized closure proposed by Kovalenko and Hirata (KH)^{26,27} — a remarkably numerically stable closure. $n \rightarrow \infty$ gives the HNC closure and its convergence issues. PSE-3 achieves a good balance between the two: it has been found to have good convergence and it gives results that approximate HNC predictions well³⁰. Due to these properties PSE-3 and other PSE- n closures have been extensively applied for a variety of charged and biomolecular systems^{32–35}.

For both HNC and its n -order expansions, the excess chemical potential of the solute at infinite dilution, μ^{ex} can be derived from the RISM solute-solvent correlation functions^{14,30,36,37}. For the HNC closure, the functional is¹⁴:

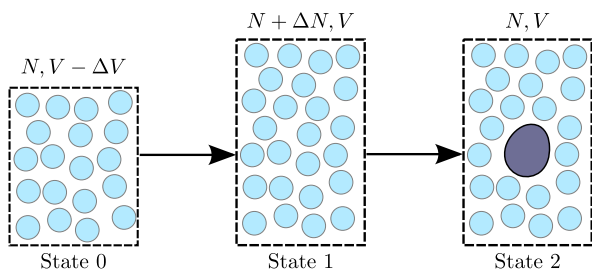
$$\mu_{HNC}^{ex} = kT \sum_{\alpha=1}^{N_{sites}} \rho_{\alpha} \int_V \left[\frac{1}{2} h_{\alpha}^2(\mathbf{r}) - c_{\alpha}(\mathbf{r}) - \frac{1}{2} c_{\alpha}(\mathbf{r}) h_{\alpha}(\mathbf{r}) \right] d\mathbf{r} \quad (3)$$

where ρ_{α} is the number density of solvent sites α . For PSE- n closure³⁰:

$$\mu_{PSE-n}^{ex} = \mu_{HNC}^{ex} - kT \sum_{\alpha=1}^{N_{sites}} \rho_{\alpha} \int_V \left[\Theta(h_{\alpha}(\mathbf{r})) (\Xi_{\alpha}(\mathbf{r}))^{n+1} / (n+1)! \right] d\mathbf{r} \quad (4)$$

where Θ is a Heaviside step function. Unfortunately, ΔG_s computed by 3D-RISM with the use of these expressions as $\Delta G_s = \mu_{RISM}^{ex}$, have a large positive bias and standard deviation when compared to available experimental data²⁵.

FIG. 1. Illustration of different reference states. The style of the figure is inspired by Figure 1 from Ref. 25. ΔV here is the partial molar volume of the solute and $\Delta N = \Delta V \rho$.



In a recent paper, Sergiievskiy *et al.* have shown that ΔG_s predictions by molecular theories like Molecular Density Functional Theory (MDFT) and 3D-RISM can be improved by introducing a correction to transform the results from the grand canonical to the isobaric-isotherm ensemble²⁵. We follow the idea from that paper in our derivation of the proper correction for ΔG_s calculated by 3D-RISM.

The main idea behind the correction is that the original 3D-RISM is defined for the grand canonical ensemble²⁸, meaning that it computes the solute excess chemical potential with regards to state 1 (figure 1; state 1 is the initial state with $N + \Delta N$ water molecules and volume V). However, experiments and MD simulations compute the free energy difference between states 2 and 0 (figure 1; state 2 is the final state with N water molecules and volume V , state 0 has N water molecules and $V - \Delta V$ volume).

A derivation of the initial state correction for the MDFT case is given in Ref. 25. Following the main result of that paper, the ΔG_s between states 2 and 0, when computed by 3D-RISM, takes the following form:

$$\Delta G_{s,ISc} = \mu_{RISM}^{ex} - \rho kT \Delta V + \rho^2 \frac{kT}{2} \hat{c}(k=0) \Delta V \quad (5)$$

where ρ is the density of the bulk solvent, ΔV stands for the partial molar volume, and $\hat{c}(k=0)$ is the value of the Fourier transformed direct correlation function of pure water at $k=0$ (it can be obtained from 1D-RISM calculation and expressed using the pure solvent isothermal compressibility: $\rho \hat{c}(k=0) = 1 - 1/(\rho kT \chi_T)$)²⁵. We will refer to this expression as the Initial State correction (ISc).

After deriving the equation 5, Sergiievskiy *et al.* argue that the ΔG_s has an extra contribution due to the change in the chemical potential of solvent (as a result of the increase in volume) equal to $\rho kT \Delta V$. They propose the following formula for computing the solvation free energy:

$$\Delta G_{s,ISc^*} = \mu_{RISM}^{ex} + \rho^2 \frac{kT}{2} \hat{c}(k=0) \Delta V \quad (6)$$

For simplicity we will refer to this expression as ISc*. This is the formula which was used in Ref. 25 to estimate the accuracy of the corrected 3D-RISM/HNC.

Equation 6 was originally developed for MDFT, not for 3D-RISM. However, the 3D-RISM is actually defined for the case of grand canonical ensemble^{14,28} meaning that the solvent potential is imposed by the outside reservoir; therefore, there should be no change in chemical potential of the solvent due to an increase in volume. Consequently, we argue that for 3D-RISM the ISc formula (equation 5) has to be used to correct the solvation free energies.

To support our statement we will present below results of an extensive benchmarking of both the ISc* (equation 6) and the ISc (equation 5) formulae for calculating solvation free energy in bulk water (hydration free energy) of various organic molecules at a range of different temperatures.

III. METHODS

To evaluate the performance of the ISc and the ISc* models at 298 K, we have used a collection of experimental hydration free energies

for small organic molecules compiled by Mobley *et al.*³⁹ The original dataset contained 504 molecules, but we found that two molecules were duplicated ("3_methyl_but_1_ene"/"3_methylbut_1_ene" and "2_methyl_but_2_ene"/"2_methylbut_2_ene"). All compounds in the dataset are non-ionic, with an average molecular mass of 113 g/mol and an average logP of 1.7 (logP was estimated using the Rdkit⁴⁰ implementation of the Wildman-Crippen algorithm⁴¹). In addition to the molecular structures and experimental data, the dataset also includes free energies obtained using the Bennett Acceptance Ratio method⁴² from MD simulations performed with the Generalized Amber Forcefield (GAFF)⁴³, AM1-BCC⁴⁴ charges, and TIP3P water⁴⁵.

The dataset with hydration free energies at varied temperatures was kindly granted to us by C.J. Cramer.^{46,47} The dataset contains measurements of hydration free energy for a variety of small, non-ionic organic molecules in a broad temperature range and was used for parametrization of the SM6T and SM8T solvation models^{46,47}. For simplicity, we will refer to this dataset by the name of the first author of the publications: the Chamberlin dataset. The experimental hydration free energy data in both datasets used here are given as $\Delta G_{hyd}^{exp} = -RT \ln c_{aq}/c_{gas}$, with concentrations in mol/L, which corresponds to the choice of standard states suggested by Ben-Naim^{48,49}. The raw version of the Chamberlin dataset contained measurements collected from various sources. To identify and eliminate possible systematic and random errors in this data, we have adopted a strategy similar to that used in the original publications (Ref. 46 and 47). We have carefully described our steps in the supporting information. The final version of the dataset contains 272 molecules with 3053 hydration free energy data points measured at temperatures between 273K and 373K; many values were averaged over multiple experimental measurements. The average molecular mass of the molecules in the dataset is 107 g/mol, average logP is 1.6. Both Chamberlin and Mobley datasets, including the structures of the molecules that we used, are provided in the supporting information⁵⁰.

The initial geometry guess for each molecule was created using the Openbabel software package^{51,52}. After that, for each molecule we performed a conformational search with the OPLS_2005 force field⁵³ using the mixed torsional/low-mode sampling method as implemented in Macromodel⁵⁴. For subsequent calculations, we used only the lowest energy conformation for each molecule. The data analysis and plotting was accomplished using the Scientific Python software stack⁵⁵⁻⁵⁸. Molecular structures were visualized using Rdkit and Marvin^{40,59}.

For all calculations we used a modified SPC/E water model (cSPC/e, see Ref. 60 for details) with oxygen partial charge of -0.8476 , $\sigma_O = 3.16572 \text{ \AA}$, $\sigma_H = 1.16572 \text{ \AA}$, $\epsilon_O = 0.1553 \text{ kcal/mol}$, $\epsilon_H = 0.01553 \text{ kcal/mol}$. For all the solutes Lennard-Jones parameters were taken from the general amber force field (GAFF)⁴³ and partial charges were generated using the AM1-BCC scheme⁴⁴.

The water susceptibility functions were generated using the DRISM program^{14,27,60} included in the AmberTools 13 package⁶¹. The set of input parameters consisted of temperature, water density and dielectric constants. The latter two parameters were obtained using interpolation functions provided in the Water Society manual⁶² (the relative uncertainty of the density is around 0.0001% and for the dielectric constant is 0.01 %). The DRISM equations were solved with tolerance set to 1×10^{-12} and grid spacing to 0.025 \AA .

All 3D-RISM calculations were performed using the rism3d.snglpnt program from AmberTools 13 package. The grid spacing was set to 0.3 \AA buffer to 30 \AA and tolerance to 1×10^{-10} . These parameters were found to converge even for rather big systems²³. All calculations were performed twice, with both HNC and PSE-3 bridge closures. The outputs obtained with different closures were not mixed, meaning that DRISM, 3D-RISM and hydration free energy evaluations were done using formulas specific to a particular closure. Integral equations were solved numerically using a modified DIIS algorithm to speed up convergence⁶³. To automate the calculation procedure we created a Python script, which we have made freely available at <https://github.com/MTS-Strathclyde/ISc>.

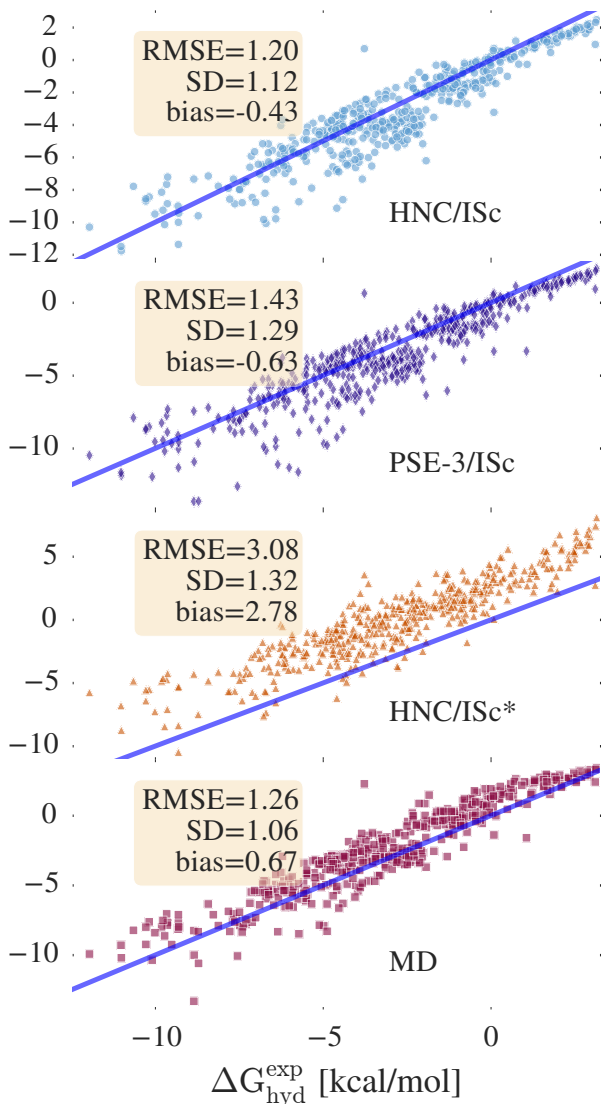
IV. RESULTS AND DISCUSSION

Figure 2 shows the accuracies of four different models on the 298 K dataset composed by Mobley *et al.* Remarkably, HNC/ISc achieves an accuracy equivalent to the MD simulations which used the same force field and partial charges (the MD data were taken from Ref. 39). While the HNC/ISc*'s standard error is not much bigger, it has a large positive bias, providing a justification for our approach to the correction of μ_{RISM}^{ex} .

Due to the above mentioned problems, calculations with HNC closure did not converge for 25 molecules (details are provided in the supporting information). On the other hand, all the PSE-3/ISc calculations converged with an accuracy only slightly worse than that for HNC/ISc. This suggest that PSE-3/ISc may be a practical alternative to HNC/ISc calculations for many systems of chemical interest.

The results predicted by both HNC/ISc and PSE-3/ISc are similar to MD: the correlation coefficient is $r \approx 0.97$ for both models (for comparison the correlation of HNC/ISc and PSE-3/ISc models with experiment was 0.93 and 0.91 respectively). The difference between 3D-RISM and MD approaches mostly originates from the difference in the signs of their biases (figure 2). This indicates that the force field quality might be one of the limiting factors for the further improvement of 3D-RISM hydration free energy quality. These issues might be addressed through the combinations of 3D-RISM with quantum chemistry⁶⁴⁻⁶⁹ or through the use of polarizable water models⁷⁰.

FIG. 2. Hydration free energies predicted by HNC/ISc, PSE-3/ISc, HNC/ISc*, and MD against experimental data from the Mobley dataset at 298 K. The blue line is a graph of the function $y=x$.



Careful examination of Figure 2 shows that ISc models perform somewhat worse than MD for hydrophilic compounds. This is most likely caused by shortcomings of the bridge closure. Essentially, the correction presented here only addresses errors due to excluded volume effects, the problems due to the exclusion of higher order bridge terms are still present. This issue might be addressed using more advanced bridges, of which the method proposed by Duh and Henderson is one such example⁷¹. This is the subject for future work.

It should be noted that the functional form of the model presented here is similar to another 3D-RISM hydration free energy correction published by Palmer *et al.* in Ref. 21: $\Delta G_{s,UC} = \mu_{RISM,KH}^{ex} + a\rho\Delta V + b$. Here

$a = -3.312$ kcal/mol and $b = 1.152$ kcal/mol are fitted coefficients for water solvent. Rearranging the ISc model into similar form gives us $a = -3.673$ kcal/mol and $b = 0$ kcal/mol for aqueous solvent at standard conditions. The correction by Palmer *et al.* gives similar accuracy for *ambient* conditions. However, ISc similarly to 3D-RISM can be applied to a system at arbitrary temperatures and pressures. Therefore, to provide additional insights into the validity of the ISc and ISc* models we performed benchmarks for each model at different temperatures.

To the best of our knowledge, there have been no reports investigating the performance of the 3D-RISM or a similar model at a wide range of temperatures for a large dataset of compounds from different chemical classes. Therefore, the benchmark provides a landmark for existing molecular theories of solvation.

FIG. 3. RMSE of hydration free energies of HNC/ISc (blue columns), PSE-3/ISc (violet columns) and HNC/ISc* (orange columns) depending on temperature. The bold lines on top of the bars are standard uncertainties (the equation is provided in the Supporting Information). The experimental data are taken from Chamberlin dataset.^{46,47}

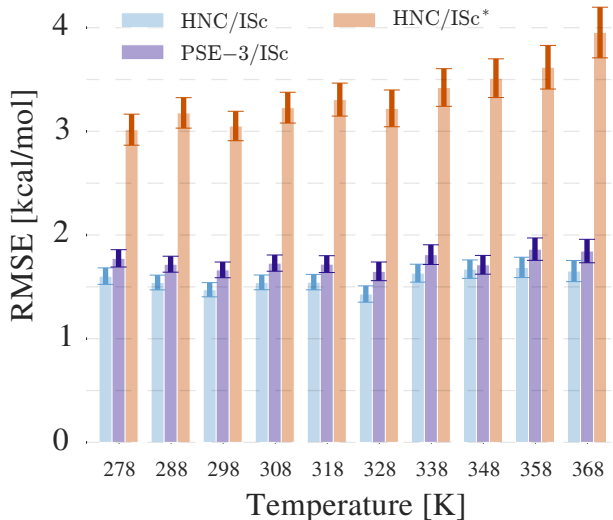


Figure 3 shows how RMSE of the PSE-3/ISc, HNC/ISc and HNC/ISc* predictions vary with temperature for the Chamberlin dataset. Each bar is computed for all points falling into ± 5 K interval. In cases where a molecule had more than one data point in a particular interval we only took into account its average error when computing the interval's RMSE. Not all intervals have the same number of molecules: there are 225 points between 293–303 K and only 130 points between 363–373 K. This is reflected by somewhat larger standard errors of RMSE towards higher temperatures where there are generally fewer points. We are only comparing the 3D-RISM models here because we were not able to find similar MD data for non-ambient temperatures.

At most temperatures, ISc models are almost twice as accurate as ISc*. Additionally, ISc* becomes less reliable towards higher temperatures, while the ISc errors remains essentially constant for all temperatures. This is another indication that supports the validity of the ISc over the ISc* model.

The error of the ISc models at 298 K is slightly higher (about 1.5 kcal/mol) for the Chamberlin dataset than for the Mobley dataset (3). This difference seems to originate from differences in the molecules as well as the sources of experimental data. It is hard to determine which data are more accurate as experimental uncertainties were not published with either dataset. As in the case of ambient temperature dataset, PSE-3/ISc results are only slightly worse than HNC/ISc ones, but have an advantage due to greater numerical stability: calculations with HNC bridge didn't converge in 121 cases (more details are provided in the supporting information).

To summarize, this letter describes a new, 3D-RISM theory based on the initial state correction (ISc). In contrast to the models that use semi-empirical corrections like the ones described in Refs 15,21,22 this model does not require additional training and/or extensive parameterisation. At the same time, the ISc models predict the hydration free energy for a wide set of organic molecules with a good accuracy both at ambient and non-ambient temperatures. In terms of accuracy the model performs best when combined with HNC; however, much better numerical stability with only a slight decrease in accuracy can be achieved with the PSE-3 closure.

ACKNOWLEDGMENTS

We are grateful to Fumio Hirata for valuable discussions about the theoretical foundations of the 3D-RISM and to Volodymyr Sergiievskiy and Daniel Borgis for helpful discussions about the corrected MDFT theory. Results were obtained using the EPSRC funded ARCHIE-WeSt High Performance Computer (www.archie-west.ac.uk). EPSRC grant no. EP/K000586/1. DSP is grateful for funding from the European Commission through a Marie Curie Intra-European Fellowship within the 7th European Community Framework Programme (FP7-PEOPLE-2010-IEF) and to the University of Strathclyde for support through its Strategic Appointment and Investment Scheme.

- ¹J. D. Thompson, C. J. Cramer, and D. G. Truhlar, "Predicting aqueous solubilities from aqueous free energies of solvation and experimental or calculated vapor pressures of pure substances," *J. Chem. Phys.* **119**, 1661–1670 (2003).
- ²D. S. Palmer, A. Llinàs, I. Morao, G. M. Day, J. M. Goodman, R. C. Glen, and J. B. O. Mitchell, "Predicting intrinsic aqueous solubility by a thermodynamic cycle," *Mol. Pharmaceutics* **5**, 266–279 (2008).
- ³D. S. Palmer, J. L. McDonagh, J. B. O. Mitchell, T. van Mourik, and M. V. Fedorov, "First-principles calculation of the intrinsic aqueous solubility of crystalline druglike molecules," *J. Chem. Theory Comput.* **8**, 3322–3337 (2012).

- ⁴R. Casanovas, J. Ortega-Castro, J. Frau, J. Donoso, and F. Muñoz, "Theoretical pKa calculations with continuum model solvents, alternative protocols to thermodynamic cycles," *Int. J. Quantum Chem.* **114**, 1350–1363 (2014).
- ⁵N. M. Garrido, A. J. Queimada, M. Jorge, E. A. Macedo, and I. G. Economou, "1-Octanol/Water partition coefficients of n-alkanes from molecular simulations of absolute solvation free energies," *J. Chem. Theory Comput.* **5**, 2436–2446 (2009).
- ⁶D. S. Palmer, J. Sørensen, B. Schiøtt, and M. V. Fedorov, "Solvent binding analysis and computational alanine scanning of the bovine Chymosin–Bovine κ -casein complex using molecular integral equation theory," *J. Chem. Theory Comput.* **9**, 5706–5717 (2013).
- ⁷N. Hansen and W. F. van Gunsteren, "Practical aspects of free-energy calculations: A review," *J. Chem. Theory Comput.* **10**, 2632–2647 (2014).
- ⁸C. Chipot and A. Pohorille, *Free Energy Calculations: Theory and Applications in Chemistry and Biology* (Springer Science & Business Media, Paris, France, 2007).
- ⁹M. R. Shirts and D. L. Mobley, "An introduction to best practices in free energy calculations," in *Biomolecular Simulations*, Methods in Molecular Biology No. 924, edited by L. Monticelli and E. Salonen (Humana Press, Totowa, NJ, 2013) pp. 271–311.
- ¹⁰D. Beglov and B. Roux, "Solvation of complex molecules in a polar liquid: An integral equation theory," *J. Chem. Phys.* **104**, 8678–8689 (1996).
- ¹¹D. Beglov and B. Roux, "Numerical-solution of the hypernetted-chain equation for a solute of arbitrary geometry in 3 dimensions," *J. Chem. Phys.* **103**, 360–364 (1995).
- ¹²A. Kovalenko and F. Hirata, "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* **103**, 7942–7957 (1999).
- ¹³A. Kovalenko and F. Hirata, "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.* **110**, 10095–10112 (1999).
- ¹⁴F. Hirata, ed., *Molecular Theory of Solvation* (Kluwer Academic Publishers, Amsterdam, Netherlands, 2003).
- ¹⁵J.-F. Truchon, B. M. Pettitt, and P. Labute, "A cavity corrected 3D-RISM functional for accurate solvation free energies," *J. Chem. Theory Comput.* **10**, 934–941 (2014).
- ¹⁶Y. Maruyama, N. Yoshida, H. Tadano, D. Takahashi, M. Sato, and F. Hirata, "Massively parallel implementation of 3D-RISM calculation with volumetric 3D-FFT," *J. Comput. Chem.* **35**, 1347–1355 (2014).
- ¹⁷V. P. Sergiievskiy and M. V. Fedorov, "3DRISM multigrid algorithm for fast solvation free energy calculations," *J. Chem. Theory Comput.* **8**, 2062–2070 (2012).
- ¹⁸Y. Harano, H. Sato, and F. Hirata, "Solvent effects on a diels-alder reaction in supercritical water: RISM-SCF study," *J. Am. Chem. Soc.* **122**, 2289–2293 (2000).
- ¹⁹T. Imai, M. Kinoshita, and F. Hirata, "Salt effect on stability and solvation structure of peptide: An integral equation study," *Bull. Chem. Soc. Jpn.* **73**, 1113–1122 (2000).
- ²⁰M. Kinoshita, Y. Okamoto, and F. Hirata, "Peptide conformations in alcohol and water: Analyses by the reference interaction site model theory," *J. Am. Chem. Soc.* **122**, 2773–2779 (2000).
- ²¹D. S. Palmer, A. I. Frolov, E. L. Ratkova, and M. V. Fedorov, "Towards a universal method for calculating hydration free energies: a 3D reference interaction site model with partial molar volume correction," *J. Phys.: Condens. Matter* **22**, 492101 (2010).
- ²²D. S. Palmer, A. I. Frolov, E. L. Ratkova, and M. V. Fedorov, "Toward a universal model to calculate the solvation thermodynamics of druglike molecules: The importance of new experimental databases," *Mol. Pharmaceutics* **8**, 1423–1429 (2011).
- ²³A. I. Frolov, E. L. Ratkova, D. S. Palmer, and M. V. Fedorov, "Hydration thermodynamics using the reference interaction site model: Speed or accuracy?" *J. Phys. Chem. B* **115**, 6011–6022 (2011).

- (2011).
- 24 E. L. Ratkova and M. V. Fedorov, "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.* **7**, 1450–1457 (2011).
 - 25 V. P. Sergiievskiy, G. Jeanmairet, M. Levesque, and D. Borgis, "Fast computation of solvation free energies with molecular density functional theory: Thermodynamic-ensemble partial molar volume corrections," *J. Phys. Chem. Lett.* **5**, 1935–1942 (2014).
 - 26 A. Kovalenko and F. Hirata, "Potentials of mean force of simple ions in ambient aqueous solution. II. solvation structure from the three-dimensional reference interaction site model approach, and comparison with simulations," *J. Chem. Phys.* **112**, 10403–10417 (2000).
 - 27 A. Kovalenko and F. Hirata, "Potentials of mean force of simple ions in ambient aqueous solution. I. three-dimensional reference interaction site model approach," *J. Chem. Phys.* **112**, 10391–10402 (2000).
 - 28 J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, 4th ed. (Academic Press, San Diego, CA, 2013).
 - 29 G. N. Chuev and M. V. Fedorov, "Wavelet algorithm for solving integral equations of molecular liquids. a test for the reference interaction site model," *J. Comput. Chem.* **25**, 1369–1377 (2004).
 - 30 S. M. Kast and T. Kloss, "Closed-form expressions of the chemical potential for integral equation closures with certain bridge functions," *J. Chem. Phys.* **129**, 236101 (2008).
 - 31 M. Kinoshita and M. Harada, "Numerical solution of the RHNC theory for water-like fluids near a macroparticle and a planar wall," *Mol. Phys.* **81**, 1473–1488 (1994).
 - 32 T. Luchko, I. S. Joung, and D. A. Case, "Integral equation theory of biomolecules and electrolytes," in *Innovations in Biomolecular Modeling and Simulations*, Vol. 1, edited by T. Schlick (Royal Society of Chemistry, London, UK, 2012) pp. 51–86.
 - 33 I. S. Joung, T. Luchko, and D. A. Case, "Simple electrolyte solutions: Comparison of DRISM and molecular dynamics results for alkali halide solutions," *J. Chem. Phys.* **138**, 044103 (2013).
 - 34 G. M. Giambaşu, T. Luchko, D. Herschlag, D. M. York, and D. A. Case, "Ion counting from explicit-solvent simulations and 3D-RISM," *Biophys. J.* **106**, 883–894 (2014).
 - 35 J. Heil, D. Tomazic, S. Egbers, and S. M. Kast, "Acidity in DMSO from the embedded cluster integral equation quantum solvation model," *J. Mol. Model.* **20**, 1–8 (2014).
 - 36 H. Saito, N. Matubayasi, K. Nishikawa, and H. Nagao, "Hydration property of globular proteins: An analysis of solvation free energy by energy representation method," *Chem. Phys. Lett.* **497**, 218–222 (2010).
 - 37 Y. Karino, M. V. Fedorov, and N. Matubayasi, "End-point calculation of solvation free energy of amino-acid analogs by molecular theories of solution," *Chem. Phys. Lett.* **496**, 351–355 (2010).
 - 38 Q. H. Du, D. Beglov, and B. Roux, "Solvation free energy of polar and nonpolar molecules in water: An extended interaction site integral equation theory in three dimensions," *J. Phys. Chem. B* **104**, 796–805 (2000).
 - 39 D. L. Mobley, C. I. Bayly, M. D. Cooper, M. R. Shirts, and K. A. Dill, "Small molecule hydration free energies in explicit solvent: An extensive test of fixed-charge atomistic simulations," *J. Chem. Theory Comput.* **5**, 350–358 (2009).
 - 40 G. Landrum, *RDKit: Open-source cheminformatics* (2013).
 - 41 S. A. Wildman and G. M. Crippen, "Prediction of physicochemical parameters by atomic contributions," *J. Chem. Inf. Comput. Sci.* **39**, 868–873 (1999).
 - 42 C. H. Bennett, "Efficient estimation of free energy differences from monte carlo data," *J. Comput. Phys.* **22**, 245–268 (1976).
 - 43 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field," *J. Comput. Chem.* **25**, 1157–1174 (2004).
 - 44 A. Jakalian, D. B. Jack, and C. I. Bayly, "Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. parameterization and validation," *J. Comput. Chem.* **23**, 1623–1641 (2002).
 - 45 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *J. Chem. Phys.* **79**, 926–935 (1983).
 - 46 A. C. Chamberlin, C. J. Cramer, and D. G. Truhlar, "Predicting aqueous free energies of solvation as functions of temperature," *J. Phys. Chem. B* **110**, 5665–5675 (2006).
 - 47 A. C. Chamberlin, C. J. Cramer, and D. G. Truhlar, "Extension of a temperature-dependent aqueous solvation model to compounds containing nitrogen, fluorine, chlorine, bromine, and sulfur," *J. Phys. Chem. B* **112**, 3024–3039 (2008).
 - 48 A. Ben-Naim, "Standard thermodynamics of transfer. uses and misuses," *J. Phys. Chem.* **82**, 792–803 (1978).
 - 49 A. Ben-Naim and Y. Marcus, "Solvation thermodynamics of non-ionic solutes," *J. Chem. Phys.* **81**, 2016–2027 (1984).
 - 50 See supplemental material at [URL will be inserted by AIP] for pdb structures as well as both experimental energies and our results for all compounds.
 - 51 N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, "Open babel: An open chemical toolbox," *J. Cheminform.* **3**, 33 (2011).
 - 52 N. M. O'Boyle, C. Morley, and G. R. Hutchison, "Pybel: a python wrapper for the OpenBabel cheminformatics toolkit," *Chem. Cent. J.* **2**, 1–5 (2008).
 - 53 J. L. Banks, H. S. Beard, Y. Cao, A. E. Cho, W. Damm, R. Farid, A. K. Felts, T. A. Halgren, D. T. Mainz, J. R. Maple, R. Murphy, D. M. Philipp, M. P. Repasky, L. Y. Zhang, B. J. Berne, R. A. Friesner, E. Gallicchio, and R. M. Levy, "Integrated modeling program, applied chemical theory (IMPACT)," *J. Comput. Chem.* **26**, 1752–1780 (2005).
 - 54 *MacroModel*, Schrodinger, LLC, New York (2013).
 - 55 F. Pérez and B. E. Granger, "IPython: a system for interactive scientific computing," *Comput. Sci. Eng.* **9**, 21–29 (2007).
 - 56 W. McKinney, "Data structures for statistical computing in python," in *Proceedings of the 9th Python in Science Conference*, edited by S. v. d. Walt and J. Millman (2010) pp. 51–56.
 - 57 S. v. d. Walt, S. C. Colbert, and G. Varoquaux, "The NumPy array: A structure for efficient numerical computation," *Comput. Sci. Eng.* **13**, 22–30 (2011).
 - 58 J. D. Hunter, "Matplotlib: A 2D graphics environment," *Comput. Sci. Eng.* **9**, 90–95 (2007).
 - 59 *Marvin*, ChemAxon Ltd., Budapest, Hungary (2014).
 - 60 T. Luchko, S. Gusarov, D. R. Roe, C. Simmerling, D. A. Case, J. Tuszynski, and A. Kovalenko, "Three-dimensional molecular theory of solvation coupled with molecular dynamics in amber," *J. Chem. Theory Comput.* **6**, 607–624 (2010).
 - 61 D. Case, T. Darden, T. Cheatham III, C. Simmerling, J. Wang, R. Duke, R. Luo, R. Walker, W. Zhang, K. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. Götz, I. Kolossváry, K. Wong, F. Paesani, J. Vanicek, R. Wolf, J. Liu, X. Wu, S. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. Roe, D. Mathews, M. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P. Kollman, *AMBER 13* (University of California, San Francisco, 2012).
 - 62 K. Daucik, "Revised supplementary release on properties of liquid water at 0.1 MPa," (2011).
 - 63 A. Kovalenko, S. Ten-No, and F. Hirata, "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.* **20**, 928–936 (1999).
 - 64 T. Kloss, J. Heil, and S. M. Kast, "Quantum chemistry in solution by combining 3D integral equation theory with a cluster embedding approach," *J. Phys. Chem. B* **112**, 4337–4343 (2008).
 - 65 N. Yoshida and F. Hirata, "A new method to determine electrostatic potential around a macromolecule in solution from molecular wave functions," *J. Comput. Chem.* **27**, 453–462 (2006).
 - 66 N. Minezawa and S. Kato, "Efficient implementation of three-dimensional reference interaction site model self-consistent-field

- method: Application to solvatochromic shift calculations,” *J. Chem. Phys.* **126**, 054511 (2007).
- ⁶⁷H. Sato, A. Kovalenko, and F. Hirata, “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.* **112**, 9463–9468 (2000).
- ⁶⁸H. Sato, N. Matubayasi, M. Nakahara, and F. Hirata, “Which carbon oxide is more soluble? ab initio study on carbon monoxide and dioxide in aqueous solution,” *Chemical Physics Letters* **323**, 257–262 (2000).
- ⁶⁹D. Casanova, S. Gusarov, A. Kovalenko, and T. Ziegler, “Evaluation of the SCF combination of KS-DFT and 3D-RISM-KH; solvation effect on conformational equilibria, tautomerization energies, and activation barriers,” *J. Chem. Theory Comput.* **3**, 458–476 (2007).
- ⁷⁰F. Hoffgaard, J. Heil, and S. M. Kast, “Three-dimensional RISM integral equation theory for polarizable solute models,” *J. Chem. Theory Comput.* **9**, 4718–4726 (2013).
- ⁷¹D.-M. Duh and D. Henderson, “Integral equation theory for Lennard-Jones fluids: The bridge function and applications to pure fluids and mixtures,” *J. Chem. Phys.* **104**, 6742–6754 (1996).
- ⁷²N. Matubayasi and M. Nakahara, “Theory of solutions in the energy representation. III. treatment of the molecular flexibility,” *J. Chem. Phys.* **119**, 9686–9702 (2003).
- ⁷³N. Matubayasi and M. Nakahara, “Theory of solutions in the energy representation. II. functional for the chemical potential,” *J. Chem. Phys.* **117**, 3605–3616 (2002).
- ⁷⁴N. Matubayasi and M. Nakahara, “Theory of solutions in the energetic representation. i. formulation,” *J. Chem. Phys.* **113**, 6070–6081 (2000).