First-principles modeling of chemistry in mixed solvents: Where to go from here?

Alex M. Maldonado (iD), Yasemin Basdogan (iD), Joshua T. Berryman (iD), Susan B. Rempe (iD), and John A. Keith (iD)

View Online   Export Citation   CrossMark

First-principles modeling of chemistry in mixed solvents: Where to go from here?

View Online   Export Citation   CrossMark

Alex M. Maldonado,[1] (iD) Yasemin Basdogan,[1] (iD) Joshua T. Berryman,[2] (iD) Susan B. Rempe,[3] (iD) and John A. Keith[1,a] (iD)

AFFILIATIONS

[1] Department of Chemical and Petroleum Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA

[2] Department of Physics and Materials Science, University of Luxembourg, L-1511 Luxembourg City, Luxembourg

[3] Center for Computational Biology and Biophysics, Sandia National Laboratories, Albuquerque, New Mexico 87185, USA

[a] Author to whom correspondence should be addressed: jakeith@pitt.edu

ABSTRACT

Mixed solvents (i.e., binary or higher order mixtures of ionic or nonionic liquids) play crucial roles in chemical syntheses, separations, and electrochemical devices because they can be tuned for specific reactions and applications. Apart from fully explicit solvation treatments that can be difficult to parameterize or computationally expensive, there is currently no well-established first-principles regimen for reliably modeling atomic-scale chemistry in mixed solvent environments. We offer our perspective on how this process could be achieved in the near future as mixed solvent systems become more explored using theoretical and computational chemistry. We first outline what makes mixed solvent systems far more complex compared to single-component solvents. An overview of current and promising techniques for modeling mixed solvent environments is provided. We focus on so-called hybrid solvation treatments such as the conductor-like screening model for real solvents and the reference interaction site model, which are far less computationally demanding than explicit simulations. We also propose that cluster-continuum approaches rooted in physically rigorous quasi-chemical theory provide a robust, yet practical, route for studying chemical processes in mixed solvents.

I. INTRODUCTION

Solvents give rise to drastic effects on chemical kinetics, thermodynamics, and reaction mechanisms.[1] Mixed solvents are appealing because their polarity and Brønsted acidity and basicity can be continuously modulated to optimize reaction outcomes.[2–7] For a recent example, tuning the concentration of hydroxyl groups in aqueous organic solvent mixtures tremendously increases rates of various acid-catalyzed reactions up to 1000 fold.[8] The synergistic effects of complicated solvent environments are clearly worth studying, but it remains unclear how to best translate practical calculation schemes used for single-component (pure) solvents (e.g., for predicting pK_as,[9] standard redox potentials,[10] and hydricities[11]) to make insightful predictions in complicated mixed solvent environments.

In this Perspective, solvation thermodynamics, which are more thoroughly discussed in other references,[12–14] will be approximated as arising mostly from electrostatic, dispersion, and configurational entropy contributions. Figure 1 graphically represents key differences between pure and mixed solvents acting on a system having a net dipole. For charged or polar solutes in polar or ionic solvents, the largest contributions to solvation free energies are electrostatic, and these terms have been well treated historically by existing theories. A significant amount of electrostatic solvation free energy is gained through the reorientation of solvent molecules to cancel the electric field of the solute. Under solvation in a pure solvent, the solvent density remains homogeneously distributed apart from heterogeneities due to formation of direct contacts with the solute [Fig. 1(a)]. However, in a mixed solvent, the relative densities of the solvent components can change, and this allows for charge migration as well as segregation of molecules having different dipole moments to result in complex double-layer structures [Fig. 1(b)].

Beyond electrostatic terms, electronic fluctuations of polarizable molecules lead to dispersion forces between the solvent and
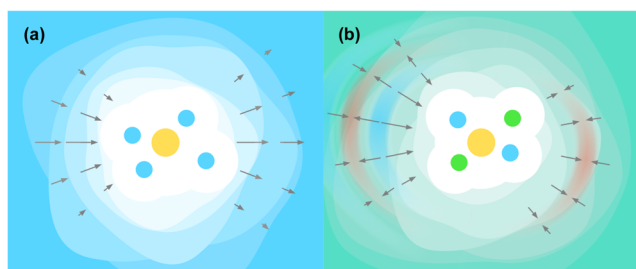
FIG. 1. (a) A molecular solute with a dipole moment (yellow sphere) may strongly interact with other molecules of the solvent (blue spheres) via electrostatic, depletion, or dispersion forces. Further away from the solute, molecular ordering is mainly driven by electrostatics (i.e., aligned dipoles, depicted as arrows) and becomes progressively weaker and more smoothly defined (blue ripples). (b) A molecular solute with a dipole moment in a mixed solvent system. Local solvent molecules (blue and green spheres) preferentially organize and segregate around the solute. The added ability of charge densities and dipoles to migrate allows for complex double-layer structures in both position and orientation, even out to the "medium" range.

the solute. Traditionally, most solvation models have treated dispersion forces as short-range interactions, but recent work has shown that electronic fluctuations can also drive long-range and highly collective interactions that are not adequately modeled as short-range two-body forces.[15,16]

Finally, flexible solute molecules may possess significant configurational entropy arising from their nuclear motion, and this can either increase or decrease (in much the same way as dispersion energies) depending on how collective degrees of freedom of the solute are coupled to solvent fluctuations. These three categories of solvation free energies are furthermore all coupled, crossing over in particular when mesoscale electrostatic effects alter the solute electronic structure.

Once the physical components of solvation free energies are accounted for, computational procedures can be used to predict reaction energies, acidity constants, standard redox potentials, and other quantities.[17,18] Using a generalized thermodynamic cycle, as shown in Scheme 1, one can compute the free energy change for
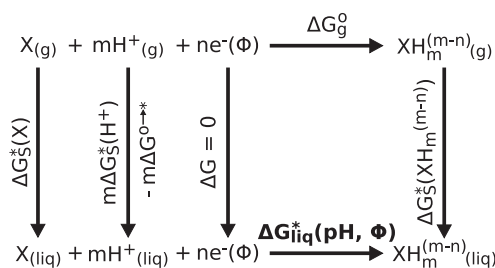
SCHEME 1. Generalized thermodynamic cycle for calculating free energy differences in different (electro)chemical environments. The top leg reflects the change in gas-phase free energy of a reaction involving a species X and different numbers of protons and electrons. Each vertical leg represents the solvation free energy of a different species so that the bottom leg will reflect the solvated free energy of that reaction. The objective is to determine the free energy of this (electro)chemical process in the liquid phase (bolded term).

a chemical process in any solvating liquid phase as the lower horizontal leg, ΔG^*_{liq}, as a function of an (electro)chemical environment (e.g., pH and an applied bias, ϕ).[18] The superscript $*$ denotes the liquid standard state of 1M. The upper horizontal leg of the cycle is the gas phase free energy difference for the same reaction (using a standard state of 1 bar, denoted by superscript $\circ$). An assortment of quantum chemical methods from Kohn–Sham density functional theory (DFT) to correlated wave function theories with corresponding ideal gas, rigid rotor, and harmonic oscillator approximations may be convenient in this setting.

The solvation free energy contributions shown by the vertical legs of this thermodynamic cycle (Scheme 1) are normally assumed to be challenging to determine precisely, as well as accurately. For instance, solvation free energies can be determined using implicit, explicit, or hybrid solvation models, but those calculated energies can be sensitive to solute geometries and simulation parameters.

Scheme 1 also requires knowledge about the proton solvation free energy in the solvent, $\Delta G^*_S(H^+)$, as well as a reference (i.e., absolute) free energy for an electron in a reference electrode interacting with that solvent, $G(e^-)$. The latter can be derived using the former, and both are known in water and some other pure solvents through semiempirical schemes.[19,20] Note that $\Delta G^*_S(H^+)$ and $G(e^-)$ are generally not known in mixed solvents. In addition, there are discrepancies in the literature due to whether the reported proton solvation free energy and its corresponding absolute electrode potential are correctly accounting for the phase potential for the solvent, as discussed later.[17,21] In the traditional derivation of Trasatti,[22] the absolute proton solvation free energy used (-1088 kJ/mol) led to an absolute reduction potential of -4.44 V for the standard hydrogen electrode. Another widely used value for the proton solvation free energy in water from Tissandier et al.,[23] -1105 kJ/mol, is a value that we believe represents a real solvation energy that includes a negative valued phase potential contribution. Kelly et al.[20] calculated a similar value to derive an absolute reduction potential of -4.28 V, a value also obtained by Isse and Gennaro[19] with $\Delta G^*_S(H^+) = -1101$ kJ/mol and Fermi–Dirac instead of Boltzmann statistics.

Based on Scheme 1, the following equation provides the complete expression to calculate the pH- and potential-dependent energy difference for any general (electro)chemical process involving m protons and/or n electron transfers to an arbitrary molecule X in an arbitrary solvent:

$$\Delta G^*_{liq}(pH, \phi) = G\left(XH_m^{(m-n)}{}_{(g)}\right) + \Delta G^*_S\left(XH_m^{(m-n)}\right)$$
$$- G\left(X_{(g)}\right) - \Delta G^*_S(X) - m\left[G\left(H^+_{(g)}\right)\right.$$
$$\left. - \Delta G^*_S(H^+) - RT\ln(10)pH + \Delta G^{\circ \to *}\right]$$
$$- n\left[G(e^-) - F\phi\right]. \quad (1)$$

Note that the pH dependence of ΔG^*_{liq} enters as linear corrections based on the Nernst relation. The dependence on ϕ is also treated as a linear correction. Combined, accounting for these corrections is sometimes referred to as the computational hydrogen electrode model.[24] In sum, Eq. (1) is useful for several straightforward calculations. The negative logarithm of the acid dissociation constant for a single proton transfer ($m = 1$, $n = 0$) is given by

$pK_a = -\Delta G^*_{\text{liq}}(\text{pH}, \phi)/[RT \ln(10)]$, where R is the gas constant and T is the temperature of the system. Likewise, the standard redox potential for an n electron transfer with no proton transfers ($m = 0$) with respect to a given reference electrode is $E° = -\Delta G^*_{\text{liq}}(\text{pH}, \phi)/[nF]$, where F is the Faraday constant. For more details about other terms, see Refs. 17 and 18.

II. CHALLENGES IN MODELING MIXED SOLVENTS

The generalized thermodynamic cycle in Scheme 1 allows for the practical determination of changes in Gibbs free energies for liquid phase reactions. Since gas-phase free energy calculations can be treated at reasonably high accuracy with appropriate quantum chemistry methods, the remaining challenges in using this cycle are the determination of solvation free energies, as well as the important terms for $\Delta G^*_S(\text{H}^+)$ and $G(e^-)$. We now discuss issues that obfuscate the determination of these values.

A. Phase potential

A frequently overlooked complexity of solvation thermodynamics arises due to the presence of two definitions for the solvation free energy: "real" and "absolute" (often called intrinsic). The "real" ΔG_S includes the reversible work for the solute to cross the vapor–liquid interface. In the case that the solvent orders at the interface, such as to create a net dipole moment perpendicular to the surface, it then follows from Gauss's theorem that there is a change in electric potential from the vapor to the liquid, called the phase potential.[25–31] Real and absolute solvation free energies are thus related by

$$\Delta G^{\text{real}}_S = \Delta G^{\text{abs}}_S + zF\chi, \quad (2)$$

where z is the charge of the solute, χ is the phase potential of the solvent, and F again is Faraday's constant. The charge z present in Eq. (2) means that there is no phase potential contribution for neutral molecules. Gauss's theorem refers to an integral over the whole surface of a closed volume; so, for cases where a solution is in a vessel (particularly in a sealed vessel, as for many industrial processes), ordering at solid-liquid interfaces may be more significant in determining the phase potential than that at the vapor-liquid interface.

In a practical sense, the phase potential of a chemically balanced reaction in a single phase will cancel so long as real or absolute solvation free energies are used consistently.[32] Being unsure whether a phase potential is included in a solvation free energy can lead to errors in properties (e.g., pK_as and redox potentials) calculated from the thermodynamic cycle in Scheme 1. Furthermore, quantifying the phase potential is difficult and lacks clear consensus.[33–35] There have been numerous classical and quantum mechanical estimates for water ranging from -1.10 V to $+3.63$ V. Investigation of other pure and mixed solvent phase potentials would require carefully crafted classical force fields or significant computational cost for molecular dynamics with quantum chemistry. There is currently little empirical or computational benchmark information for the phase potentials of mixed solvents with air, indicating a need for the further work in this field. In the absence of accurate phase potential calculations or measurements, the best (but unsatisfactory) approximation is to refer to any known solvents hypothesized to have a comparable surface structure to the system at hand.

B. Preferential solvation

Preferential solvation is the local concentration variation in a mixed solvent that occurs when one chemical species has a different affinity for the solute than others. In the example of Fig. 1(b), this situation leads to formation of alternating layers. For complex species or for strong concentration gradients, this phenomenon can become highly intractable, even leading to local demixing and the formation of droplets or other structures. In the special case of monatomic ions around charged or polar solutes, Poisson–Boltzmann–Debye–Hückel theory is available to rewrite the Coulomb electrostatic potential with an additional ion screening term, solved in spherical symmetry to give $\phi \propto e^{\kappa/r}/\epsilon r$ (where ϵ is the dielectric constant and $1/\kappa$ is the Debye length over which screening by mobile charges becomes significant), rather than $\phi \propto 1/\epsilon r$ as in the absence of salt.[36] Poisson–Boltzmann–Debye–Hückel theory is powerful, but it neglects shape effects and nonelectrostatic forces, thus being mainly useful for monatomic cations, and mainly far from the solute surface. Additions to the theory to treat the modification of ion density in a short range by finite radius of ions and the presence of nonelectrostatic interactions exist and can be highly effective.[37] However, the theory remains more useful for long-range electrostatics but less so for steric and dispersion effects, which are usually in the short range.

Other theoretical frameworks provide a preferential solvation parameter on a general basis for a solute X in a mixture of i and j, δ_{Xi}. This parameter is defined as the difference between the local and bulk mole fractions (x) of a solvent component i around the solute X, $\delta_{Xi} = x^L_{Xi} - x_i$. For example, Kirkwood–Buff (KB) theory relates preferential solvation to integrals of the radial distribution function in the form of a KB integral (KBI). Such integrals usefully distill the affinity between X and i to a scalar greater than one for two species that associate, or less than one for two species that repel each other.[38,39] Inferring δ_{Xi} can be achieved by using the concentration dependence of experimental data such as isothermal compressibility, partial molar volumes, or excess molar Gibbs free energy of mixing. Additionally, the standard molar Gibbs free energy of transfer (ΔG_{tr}) of the solute from a reference solvent i to a mixture of i and j is required to obtain the KBI via the inversion of KB theory.[40] These properties and their derivatives must be determined with high accuracy, such that experiments may be difficult or expensive,[41] so the alternative of molecular dynamics is sometimes used to determine the KBI. Convergence of integrals over correlation functions using molecular dynamics is, however, difficult due the large weight given to long-range parts of the correlation function (which are also the slowest to sample),[42,43] ensemble effects, and systematic force field errors that may be present in medium and long ranges. The preferential solvation parameter is directly dependent on the derivatives of these properties, thus motivating the need for highly accurate parameterizations.

The quasi-lattice quasi-chemical (QLQC) method is another way to predict δ_{Xi}.[44–46] The QLQC method also requires information typically acquired through experiments such as ΔG_{tr} and excess molar Gibbs free energy of mixing of components i and j at

equimolar composition. Additionally, the QLQC method requires the selection of a lattice parameter, Z, which defines the number of neighbors around the solute and every solvent molecule. Z is conceptually related to the coordination number of the solute (and solvent molecules), but it mostly serves as a fitting parameter for the excess Gibbs energy of mixing curve. This adds emphasis to quality experimental data and can make *a priori* prediction of δ_{Xi} difficult. The QLQC method typically uses the same Z for all compositions of the same mixed solvent; this would be valid if the coordination number is independent of composition, which is not always the case. There are ways to have concentration dependence in Z, but this is generally avoided because it greatly complicates calculations and reduces intuitive understanding. In addition, the assumption of ideal mixing of the neighboring molecules in the quasi-lattice sites is inherently incompatible with preferential solvation. With all of these characteristics, the QLQC method is typically restricted to small solutes with sufficient experimental data.[47]

In summary, the above models and others[48–50] are useful and powerful for predicting preferential solvation effects for various solutes in mixed solvents. However, each of them has limitations due to computational cost, narrow range of well-treated cases, or requirement for specific empirical parameters. One might be tempted to average properties of the pure solvent components linearly to obtain a reasonable approximation to a mixed solvent system, as discussed in some textbooks on the subject.[46,51] This approach is enticing because it avoids the effort of predicting preferential solvation and seems plausible, but in general, mixed solvent properties exhibit nonlinearities and solute-solvent properties are dominated by the local solvent structure. Figure 2 shows the Gibbs free energy required to transfer Na^+ or CH_3COO^- from pure water to a mixture of water/acetonitrile at varying mole fractions.[52,53] Both ions deviate from the linear interpolation line between pure water and pure acetonitrile. Na^+ even has favorable transfer Gibbs free energy at acetonitrile mole fractions ranging from 10% to 50% compared to transfer from pure water to pure acetonitrile. Thus, simple averaging schemes not only are sometimes unphysical approximations but may also produce significant errors that arise from preferential solvation effects.
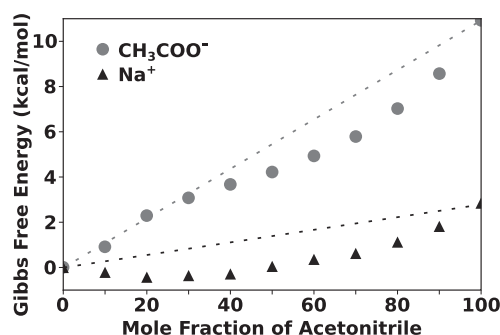

FIG. 2: Gibbs free energy of transfer of Na^+ (triangles) and CH_3COO^- (circles) from water to the water/acetonitrile mixture of varying concentrations. Energies exhibit concentration-dependent nonlinearities from preferential solvation and other cosolvent effects.

III. MODELING MIXED SOLVENTS

Explicit solvation methods involve placing the solute(s) in a large box surrounded by solvent molecules, where forces on each atom are calculated to determine the trajectories for the entire system. This straightforward technique over time has become a ubiquitous treatment for mixed solvent systems. Evaluation of the forces involves classical force fields, quantum chemical methods, and combinations of both. However, each method is dependent either on accurate parameterizations or on computationally expensive molecular dynamics with quantum chemistry. Useful information about mixed solvent systems is available from explicit modeling, but other less costly methods should be considered.

An implicit solvation model is an alternative to explicit methods. In this case, the solvent is treated as a dielectric continuum, possibly with electrostatic screening by mobile ions. Simplifying the solvent to a homogeneous medium requires extensive experimental data for significant parameterization. Furthermore, implicit solvation models calculate only a single parameter per lattice point (the electrostatic potential or its gradient), meaning that complex ordering of co-solvents, often important near the solute, cannot be captured. Forfeiting the capability to account for preferential solvation makes mixed solvents challenging for implicit solvation models, especially for polar or ionic chemical species. However, some implicit solvation models [e.g., the Solvation Model based on Density (SMD) and the Integral Equation Formalism (IEF)] have seen success with ionic liquids, but so far applications have been limited to small, neutral molecules.[54,121]

Improved theoretical foundations are needed to physically and accurately allow for predictions of solutes in mixed solvent environments without the high computational cost of explicit quantum solvation. Two such models, COnductor-like Screening MOdel for Real Solvents (COSMO-RS) and reference interaction site model (RISM), are briefly summarized as they have been demonstrated to model mixed solvents with reasonable accuracy and computational cost.

A. COSMO-RS

The COnductor-like Screening MOdel for Real Solvents (COSMO-RS) is classified as a hybrid solvation model.[55–59] Every COSMO-RS calculation starts from the COSMO implicit solvation model cavity with an infinite dielectric (perfect conductor) reference state ($\epsilon = \infty$). The local interaction energy, $E_{int}(A, B)$, of two molecules contacting each other in a liquid is quantified in COSMO-RS from the COSMO polarization charge densities, σ and σ', from their respective COSMO surfaces. COSMO-RS sums nonideal screening from unbalanced σ and σ' contributions (so-called electrostatic misfits) and hydrogen bonding by using surface segments with a_{cont} contact surface area to account for $E_{int}(A, B)$.

By summing all energy corrections over all molecular surface contacts, COSMO-RS obtains a total energy correction for that specific molecular configuration. This energy correction then needs to be expanded into an ensemble to represent a liquid. Instead of employing molecular dynamics or Monte Carlo procedures, a simpler statistical thermodynamics model of an ensemble of independently pairwise interacting surface segments is used.

In COSMO-RS, the chemical potential of a unit surface segment of kind σ in any pure or mixed solvent S is

$$\mu_S(\sigma) = \frac{-k_B T}{a_{cont}} \ln \int p_S(\sigma) \exp[-\beta a_{cont}(e(\sigma, \sigma') - \mu_S(\sigma'))] \, d\sigma, \quad (3)$$

where $\beta = 1/k_B T$, k_B is the Boltzmann constant, and $e(\sigma, \sigma')$ is the surface-specific interaction energy of polarization charges σ and σ'. $p_S(\sigma)$ is the sigma profile of the solvent or mixture S, which is essentially the reduction of the 3D σ-potential information into a histogram that conveys the probability of finding a surface segment of kind σ.

COSMO-RS can use this ensemble of surface segment chemical potentials to calculate the chemical potential of the solute X in the solvent S with

$$\mu_S^X = \int p^X(\sigma)\mu_S(\sigma) \, d\sigma + k_B T \ln \gamma_{comb}(X, S) + k_B T \ln x_X. \quad (4)$$

Note that μ_S^X is the same as $\Delta G_S^*(X)$ (i.e., the solvation free energy for a solute X mentioned earlier) but on a per mole basis. The first term in Eq. (4) represents the solute-solvent (surface segment chemical potentials) interactions described by COSMO-RS in terms of σ. The second and third terms represent the combinatorial contribution of the chemical potential arising from the dependencies of solute size and concentration, respectively.

Overall, COSMO-RS is a valuable tool for chemical engineering thermodynamics.[58] The COSMO-RS scheme significantly simplifies the problem of lengthy explicit simulations and produces solvation free energies in pure and mixed solvents. A self-consistent generalization of COSMO-RS, Direct COSMO-RS (DCOSMO-RS), has been implemented in several quantum chemistry packages.[60]

While COSMO-RS has several significant improvements over COSMO and other implicit solvation models, it still experiences some shortcomings. COSMO-RS also requires parameterizations for nonelectrostatic contributions from cavitation, dispersion, and repulsion,[61] but robust parameterizations in turn require extensive databases. Thus, approximate parameters are generally used in practice even though they might not be suitable for highly accurate computational investigations. COSMO-RS was originally developed and parameterized for the neutral solute and solvents and incorrectly assigned misfit charges while modeling charged solutes.[62] However, recent developments have shown promising success with ionic liquids[63] and electrolytes.[64] In addition, an extension called COSMO-RSC was developed to relieve some of the error associated with highly localized charges; however, these corrections are not consistently implemented in quantum chemistry packages and thus should be used with caution.[21,65]

B. RISM

The reference interaction site model (RISM) is another hybrid method, based on the integral equation theory of liquids. It was first developed in 1972 by Chandler and Anderson,[66] but it has gone through several iterations to make it an intriguing solvation model for the study of future problems. Extensive literature on the RISM already exists,[67–71] so only a brief summary is given here.

In the 1D-RISM, a single radial distribution for each solvent site around the solute is determined. In the so-called 3D-RISM, a three-dimensional density distribution $g_\gamma(\mathbf{r})$ for each interaction site (γ) of

each solvent species present is determined. A value of $g_\gamma(\mathbf{r})$ greater or less than one shows solvent site enrichment or depletion, respectively. Thermodynamic quantities, such as entropy, enthalpy, and pressure, are then determined as integrals (single, double, or higher) over the 3D density fields of the various interacting sites.

An advantage of working with 3D density maps is that distributions of specific atoms or functional groups of the solvent(s) in relation to the solute can be plotted directly. Unlike implicit solvation models, the 3D-RISM directly accounts for nontrivial ordering in the solvent, such as electric double-layers or even more complex density variations, as illustrated with a peptide fibril structure[72] in Fig. 3.

The 3D-RISM integral equation method is a set of approximate solutions for the Ornstein–Zernike equation that writes the "total" (many-body) correlation at $\mathbf{r}$ from a test species or site to a second site γ $[h_\gamma(\mathbf{r})]$ as a sum over volume integrals of "direct" (two-body) correlations to sites α (c_α) times the total correlations between sites α and γ ($h_{\alpha\gamma}$),

$$h_\gamma(\mathbf{r}) = c_\gamma(\mathbf{r}) + \sum_\alpha \rho_\alpha \int c_\alpha(\mathbf{r} - \mathbf{r}')h_{\alpha\gamma}(\mathbf{r}') \, d\mathbf{r}'. \quad (5)$$

$h_\gamma(\mathbf{r})$ of the interaction site γ is then related to the density distribution by $g_\gamma(\mathbf{r}) = h_\gamma(\mathbf{r}) + 1$. Equation (5) has indices γ and α that enumerate over all sites on all solvent species. The number density of the solvent site α is ρ_α. The total correlation function $h_{\alpha\gamma}(\mathbf{r})$ appears as an unknown one on both sides of the equation, making iterative solution (and further constraint of the relationship between c and h) necessary.

The 3D-RISM connects the 3D total correlation $h_{\alpha\gamma}(\mathbf{r})$, an emergent many-body property, and the direct two-body correlation functions $c_{\alpha\gamma}(\mathbf{r})$. The two-body correlation is directly determined by the susceptibility and can be calculated to a good approximation for small fluctuations directly from the pairwise interaction
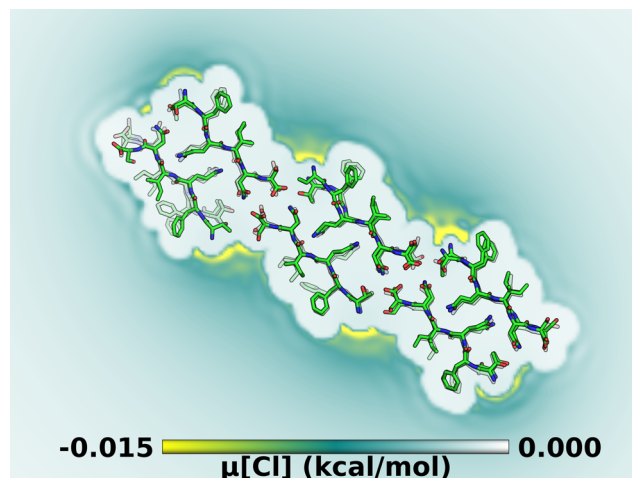

FIG. 3. Output from a calculation using the tool **3drism.snglpnt**,[75] showing a cross section cutting the long axis of an extended peptide fibril in 10 mM HCl. μ[Cl] is the chemical potential field for chloride ions, showing a clear structure even out to the third shell.

potential $-\beta u_{\gamma\alpha}(\mathbf{r})$. Such two-body correlations are also observable with scattering experiments made at low concentrations such that many-body effects are limited. Conversely, the theory can also find intermolecular potentials based on $h_{\alpha\gamma}(\mathbf{r})$ from scattering experiments at high concentrations.

The variation between 3D-RISM implementations and much of the art in applying this method arises from the choice of a closure relation, which is necessary to render Eq. (5) tractable by fixing $c_\gamma(\mathbf{r})$ in relation to $h_\gamma(\mathbf{r})$ and also to control the entry of system-specific information via $u(\mathbf{r})$. There are several closure relations, but two popular ones are the hypernetted chain (HNC)[73] approximation and Kovalenko–Hirata (KH).[74] The solvation free energy can then be found, for example, by using the KH closure resulting in

$$\mu_S^X = \frac{\rho}{\beta} \sum_\gamma \int \frac{1}{2}[h_\gamma(\mathbf{r})]^2 \Theta[-h_\gamma(\mathbf{r})] - c_\gamma(\mathbf{r}) - \frac{1}{2}h_\gamma(\mathbf{r})c_\gamma(\mathbf{r}) \, d\mathbf{r}, \quad (6)$$

where Θ is the Heaviside step function originating from the KH closure that ensures the term h^2 is only in effect in regions of solvent site depletion.

By this point, it should be clear that the 3D-RISM can be highly effective for analyzing the role of mixed solvents.[76–78] The computational expense of the 3D-RISM depends on which closure relation is chosen, as well as the parameters (e.g., Lennard-Jones) of the given system. As the solvent increases in complexity (i.e., the number of co-solvents or number of sites treated per solvent), the number of pair correlations and therefore cost of a self-consistency iteration rise as the square of the number of sites. Furthermore, the number of density field grids needed increases per site and thus requires significant computer memory resources. The 3D-RISM is therefore quite elegant but also very challenging for complex and flexible solvents. These challenges can be reduced if complicated co-solvents are treated as being present only in the low concentration limit so that solvent-solvent site interactions can be ignored within the problematic species.[77] The required fineness of the grid increases with the sharpness of concentration gradients; therefore, small or strongly interacting solvent/co-solvent sites can dramatically increase the computational cost. The generalization of the method to adaptive grids would appear to be a potential avenue for progress. Strong interactions and high concentrations also increase the importance of three-body and higher terms in the mapping from pair to total correlations, thus increasing the likelihood of slow or failed convergence depending on the closure.

Because the 3D-RISM can use pairwise interaction potentials as an input, the user is at liberty to supply the input data at arbitrary levels of quantum mechanical accuracy, although of course many-body electronic interactions cannot enter via the two-body parameterization process. With all of these technical aspects aside, the 3D-RISM is convenient to use with conventional DFT calculations, and it also has overlapping characteristics as an algorithm, so it has been integrated into some quantum chemistry packages such as Amsterdam Density Functional (ADF).[79]

IV. CLUSTER CONTINUUM

Another technique commonly used to model solvents is cluster-continuum or mixed implicit/explicit solvation modeling. Explicit solvent molecules are added around the solute and then this cluster is placed in an implicit solvation model, such that the blue region described as a "large volume with weak order" in Fig. 1(a) is treated as a continuum dielectric, while the inner region is modeled classically or via quantum chemistry. This procedure is advantageous for several reasons, including balanced computational costs and retention of local solvent environment characteristics.

Several challenges still remain, and there is a lack of widely accepted best practices for cluster-continuum modeling. For example, there is no clear consensus on choosing the number and location of the explicit solvent molecules. Often, chemical intuition has guided the placement of solvent molecules, but this is limited by *a priori* understanding of the chemical process.[80] In this section, we present our interpretation for modeling chemistry in a complicated, mixed solvent environment.

A. Quasi-chemical theory

Partitioning the contributions of the inner and outer solvation shells around a solute is fundamental in these cluster-continuum solvation schemes. Quasi-chemical theory (QCT) represents the most robust method to partition the system and calculate the solvation free energy of any solute. Only a brief introduction will be given here as there is already substantial literature on the subject.[27,81–84]

QCT is rooted in the potential-distribution theorem,[85] which provides an expression for the excess chemical potential (solvation free energy) of the solute X, $\mu_X^{(ex)}$, as a sum of free energy contributions,

$$\mu_X^{(ex)} = -k_B T \ln\left[p_X^{(0)}(n)\right] + k_B T \ln[p_X(n)] + \mu_X^{(ex)}(n), \quad (7)$$

where $p_X(n)$ is the probability of observing n ligands (solvent molecules) in the inner shell and $\mu_X^{(ex)}(n)$ is the excess chemical potential of X having exactly n ligands in the inner shell. $p_X^{(0)}(n)$ represents the probability that n ligands occupy the inner shell with no solute-solvent interactions. In other words, this quantity describes the probability of ligands occupying the designated inner-shell volume without X actually being present.

Within the QCT framework, the inner solvation shell is thought of as a chemical equilibrium reaction concerning the association of ligands (L) with the solute,

$$X + nL \rightleftharpoons XL_n. \quad (8)$$

This reaction in Eq. (8) is described by a chemical equilibrium ratio defined by $K_n \rho_L^n = p_X(n)/p_X^{(0)}(n=0)$, or the probability ratio of having n ligands to zero ligands within the solvation shell. Substituting this definition of K_n into Eq. (7) evaluated for $n = 0$, and expressing K_n in terms of an ideal gas-phase chemical equilibrium constant, $K_n^{(0)}$, leads to the cluster QCT formula[82]

$$\mu_X^{(ex)} = -k_B T \ln\left[K_n^{(0)}\rho_L^n\right] + k_B T \ln[p_X(n)] + \left(\mu_{XL_n}^{(ex)} - n\mu_L^{(ex)}\right). \quad (9)$$

Here, the excess chemical potential is broken into three terms. The first being concerned with the ligand-association chemical reaction in an ideal gas phase is denoted by the (0) superscript. An advantage of the QCT formulation is that calculation of $K_n^{(0)}$ may be accomplished with gas-phase electronic structure calculations for the free energies of the solute, individual ligand, and solute-ligand

cluster. The overall availability of ligands for the reaction is represented by the bulk density of the ligand, ρ_L. This association reaction is designated to occur within a volume chosen to represent the inner shell.

The second term accounts for the thermal fluctuations of having n ligands occupying the inner shell, defined by considering the probability of observing that particular case, $p_X(n)$. Evaluating this probability involves molecular dynamics, usually with quantum chemistry, with adequate sampling of the solute surrounded by the solvent. In some circumstances, classical force fields can be used, but then results become dependent on the quality of parameterization.[86]

The outer shell (bulk) contribution to the solvation free energy of X is captured in the third term. Essentially, it is the free energy contribution of solvating the solute-ligand complex and removing the ligands from the outer solvation environment. A standard implicit solvation model, while an approximation, has effectively evaluated $\mu_{XL_n}^{(ex)}$ and $n\mu_L^{(ex)}$ in previous applications.[27,87–98]

In QCT, selection of the inner solvation shell is typically done manually. While any shape can be used, a sphere with the radius λ centered around X is commonly used. The selection of λ is generally chosen to capture the innermost-solvation shell defined by a neighborship analysis of solvent molecules that contribute to the first radial distribution function peak;[27,99,100] however, any choice of λ is valid in QCT. Ultimately, the choice of λ dictates the probabilities of observing n ligands in the inner shell, and any case of n ligands for the inner solvation shell can be evaluated in QCT.

Some care toward selecting λ and n can deliver tremendous computational feasibility and minimize errors for several reasons. Implicit solvation models depend on the definition of the cavity to calculate solvation free energies, which are sensitive to the radii being used. When a complete solvation shell is defined, the free energy of solvation by the implicit solvation model is less sensitive to specific radius definitions. This trend suggests choosing λ that captures complete inner-shell occupancies for convenient error cancellation. While any number of complete solvation shells could be treated, the simplest inner shell volume contains only the nearest neighboring ligands that are continuously interacting with the solute. This "no split occupancies rule" means λ should be selected by a neighborship decomposition analysis,[99] where distant neighbors that encompass many shells (multimodal) are excluded.[27,92,96,100]

As previously mentioned, $p_X(n)$ must be determined by the observation frequency. Choosing n that is rarely observed would require long simulation times.[93] When molecular dynamics with quantum chemistry is used, the higher computational cost motivates choosing the most probable n within λ for the solute, $\bar{n}$. Applying QCT to the case of $\bar{n}$ would have $p_X(n)$ near one, so the term can be dropped altogether. The case of $\bar{n}$ with QCT can reduce the computational cost, but it is not always known. While $\bar{n}$ can be determined from explicit simulations and neighborship analyses, it can become computationally impractical for large solutes without vetted, highly accurate force fields. This makes direct implementation into computational chemistry challenging.

QCT has produced predictions of solvation free energies in excellent agreement with experimental values for hydration of cations and anions,[27,87–93,95–98] as well as other molecules,[101,102] and of free energies of ion binding to proteins and solvation by nonaqueous solvents.[27,83,86,94,95,100,103–106] In addition, QCT predictions of pK_{as} and redox potentials agree well with experimental values.[107,108] While QCT is completely generalized for any solute and solvent, its application to mixed solvents has not yet been published. Nevertheless, these characteristics of QCT demonstrate its robustness and theoretical rigor for calculating absolute solvation free energies.

The advantage of QCT is the separation of solvation free energies into components from inner-shell and outer-shell solvent molecules, permitting exploitation of the known chemical physics analyses of clusters. Issues that can be treated in cluster analyses include proper overlap repulsions, polarizability, charge transfer, London dispersion interactions, n-body ligand interactions generally, distortion of flexible ligands, and zero-point motion of the cluster. Large displacements in the vibrational motions of the clusters can be treated with molecular dynamics, as done recently for anions using quantum chemical simulations, and earlier for hydrogen using classical force field simulations.[98,102] Current implementations of QCT require significant effort from manual evaluation of terms, with molecular dynamics simulations and neighborship analyses used to guide selection of inner and outer solvation shells and determine $\bar{n}$ for computational feasibility and to minimize errors.

### B. Local solvation motifs using machine learning

When modeling chemical processes (e.g., reaction mechanisms with several intermediates), QCT may become impractical due to the high computational cost of molecular dynamics or requirement of accurate force field parameters for all structures (i.e., reactants, products, intermediates, and transition states) to determine $p_X(n)$ or $\bar{n}$. However, theoretical and conceptual foundations of QCT are valuable for developing future solvation schemes. Here, we discuss a novel, automatable procedure for studying solvent effects and computing solvation free energies.[109] The key feature is the integration of the QCT framework twice (ligand free energy and ion solvation free energy) into a standard thermodynamic cycle in combination with a machine learning analysis of local solvent environments. This dual-QCT approach achieves additional error cancellation by balancing QCT terms in such a way that evaluating $p_X(n)$ or seeking $\bar{n}$ and using harmonic approximations are unnecessary. Additional advantages are automation and eliminating potentially cost-prohibitive molecular dynamics with quantum chemistry and neighborship analyses.

The foundation of this solvation technique is the cluster-continuum approach represented by the thermodynamic cycle, commonly referred to as the cluster cycle, shown in Scheme 2.[110] Conceptually, a solvent cluster with n molecules is desolvated, $-\Delta G_S^*((L)_n)$, a solute X binds to the solvent cluster in the gas phase, $\Delta G_{g,\,bind}^\circ$, and then the solute-solvent cluster is solvated, $\Delta G_S^*(X(L)_n)$. The solvation free energy of the solute X can be calculated with

$$\Delta G_S^*(X) = \Delta G_S^*(X(L)_n) - \Delta G_S^*((L)_n) + \Delta G_{g,\,bind}^\circ$$
$$- \Delta G^{\circ \longrightarrow *} - RT \ln([L]/n). \quad (10)$$

There is a standard state correction, $\Delta G^{\circ \longrightarrow *}$, remaining because of the different numbers of species on the left-hand and right-hand sides of the cluster cycle. In addition, when desolvating the solvent
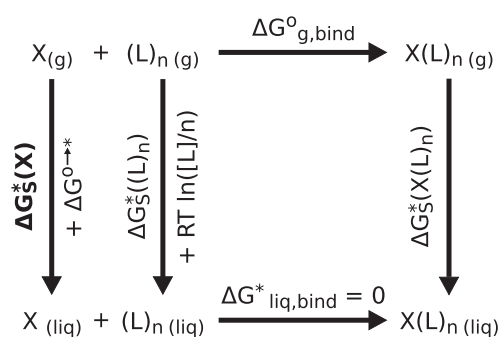
SCHEME 2. A thermodynamic cycle of solvating a solute, X, by using a preformed solvent cluster containing n solvent ligands, L. The top horizontal leg represents binding the free energy of binding X to a solvent cluster in the gas phase. Vertical legs represent the required free energy to solvate each chemical species from the gas to the liquid phase with the appropriate state corrections. The bottom leg represents the same binding reaction, but in the liquid phase that requires no free energy. We are interested in determining the solvation free energy of the solute (bolded term).

cluster, there is an additional state correction from 1M (*) to the molar concentration of the solvent, [L], in the term $RT \ln([L]/n)$ that is the density term in QCT. The solute solvation free energy problem is consequently transformed into the calculation of solvent and solute-solvent cluster solvation free energies. When QCT [Eq. (9)] is applied to $\Delta G_S^*((L)_n)$ and $\Delta G_S^*(X(L)_n)$ with any n, there are two inner shell contributions: $K_n^{(0)} \rho_L^n$ and $p(n)$. Because the two clusters are similar in size, error cancellation from the cluster cycle renders the need for a precise understanding of $\tilde{n}$ or evaluation of $p(n)$ less necessary. The remainder of this section will detail how local solvent environment analysis using machine learning algorithms guides quantification and placement of solvent molecules around a solute.

The central idea is to generate many candidate clusters of different structures and a number of solvent molecules (sizes). Then, clusters that satisfy specific criteria are used in the cluster cycle to calculate the solute solvation free energy. In past applications of the cluster cycle, solvent molecules are manually placed or carved out of an explicit simulation. This route proves to be tedious or computationally inefficient. Alternatively, an algorithm can be used to automatically generate the prospective clusters. One such example is ABCluster, which employs the artificial bee colony algorithm to find global and local minima of atomic or molecular clusters.[111,112] For each size, thousands of clusters can be rapidly generated and initially ranked according to their energy with classical force fields. Then, the lowest (e.g., five) can be further optimized at 0 K with quantum chemical methods such as DFT. To achieve an adequate representation of the local solvent environment with this multistep filtering procedure, the Boltzmann-weighted average of same sized clusters is taken. Note that the cluster cycle requires solvation free energies for solvent and solute-solvent clusters, so the above procedure is performed both in the absence and in the presence of the solute.

This still leaves the main challenge of determining the appropriate cluster size unsolved. One could compare to experimental data to solve this problem; however, this would not be different from current practices of *ad hoc* selection of the number of solvent molecules. Furthermore, having experimental solvation energies for any and all chemical species and environments is an unfeasible endeavor; there need to be separate selection criteria independent of experimental data.

As mentioned before, the local solvent environment is the largest contributor to solvation energetics, so this is a natural choice for selection criteria. To avoid requiring *a priori* information about the local solvent environment, a separate technique can be used to compare clusters of different sizes. Comparing the structures of molecules, or clusters, is not novel and is commonly done through connectivity and distances as in root mean square displacement (RMSD). Calculating RMSDs, however, requires manual alignment of the systems being compared that becomes impossible when comparing clusters containing different numbers of atoms. Over the last few years, several fingerprint functions have been used to characterize the environment of atoms, but one function that fits all does not apply here. Each is designed to differentiate between specific system characteristics. For example, some fingerprint functions characterize a system by elemental characteristics or bond angles which would not be able to differentiate local solvent environments. Hence, selecting an appropriate fingerprint function is paramount for the critical evaluation of these microsolvated clusters.

One promising algorithm is the smooth overlap of atomic positions (SOAP) kernel that enables comparison between clusters that is invariant to translations, rotations, and permutations.[113,114] The SOAP kernel evaluates the local environment by placing a Gaussian function centered on each atom. Then, the SOAP gathers information from where there is overlap of Gaussian functions (local atomic neighbor densities) and reports high-dimensional pair-similarity data. Essentially, this quantifies the similarities and differences of the local solvent environment of different clusters.

After employing the SOAP kernel, one has detailed information about the local solvent environments but needs an intuitive way to compare them. A data dimensionality reduction and visualization scheme called the sketch-map can be used to compare and visualize the high-dimensional pairwise data from the SOAP.[115,116] The sketch-map algorithm optimizes a nonlinear objective function that finds a metric, typically Euclidean distance, which best describes the high-dimensional SOAP metrics. Basically, it takes the SOAP data and converts them to a 2D space for human readability. There is an abundance of data that can be gleaned from a sketch-map analysis. One potentially confusing aspect is the axes that have no physical meaning; they could even be flipped without changing the information from the SOAP. Conclusions about the similarity of the local solvent environment can only be made qualitatively from the distances between points. In other words, points that are near each other (clustered) represent similar local solvent environments.

An example of a sketch-map analysis on the solvation free energy of Na$^+$ in water is shown in Fig. 4(a).[109] The dots are color coded based on the number of solvent molecules (in this case water) included in the cluster. To analyze the sketch-map, the proximity of dots is examined. For example, the dark slate blue dots (all having four water molecules) in the upper middle of the sketch-map have some space between them and are not very close to each other.
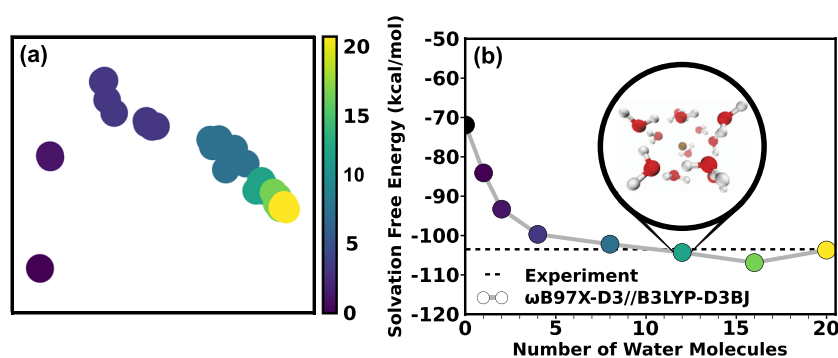
FIG. 4. (a) SOAP/sketch-map representation of microsolvated clusters with various numbers of water ligands demonstrating optimal clusters at 12 water molecules. The color bar represents the number of water molecules in the cluster. (b) Solvation free energies of Na$^+$ with the variable number of water ligands from B3LYP-D3BJ/def2-SVP geometries and ωB97X-D3/def2-TZVP calculations.[109]

This means in the five local minima, clusters generated with four water molecules have different local solvent environments. This communicates that the Gaussian functions placed on the atoms in each cluster overlapped in different amounts and locations representing dissimilarity in the SOAP. To the right in the sketch-map, the dark green dots cluster more for the eight water molecule case signifying the structures and local solvent environment begin to look similar.

The sketch-map places the 12 water clusters very close to each other and even overlapping with the 16-water molecule case. This has important significance for determining the correct cluster size for the cluster cycle. When additional solvent molecules are added, they start populating outer solvation shells and are not captured within the solute's Gaussian function. Namely, the local solvent environment starts to look the same. Once this occurs, the appropriate cluster size (12 water molecules for Na$^+$) is used with the cluster cycle to calculate the solvation free energy to avoid errors from the implicit solvation model's sensitivity to radius definitions due to incomplete solvation shells. The solvation free energies calculated from the cluster cycle using *n* water molecules for Na$^+$ are shown in Fig. 4(b). Starting from zero water molecules (around -30 kcal/mol error), the cluster cycle solvation energies begin to approach the experimental value of -101.3 kcal/mol.[23] Using the sketch-map selection criteria of 12 water molecules provided the most accurate solvation free energy.

The solvation technique discussed here attempts to capture the following three criteria in an automatable fashion. First, information of the local solvent environment needs to be retained. Preferential solvation must be accounted for to ensure accurate solvation energetics in mixed solvents. Second, explicit simulations should not be required. While substantial information can be gained, the computational cost of explicit simulations and/or their reliance on parameterization prohibits their routine use. Third, no *a priori* information about the system should be required. For example, knowing the coordination number of the solute drastically reduces the number of applicable systems or demands explicit simulations.

This approach has already been successfully tested in multiple systems. Basdogan and Keith were able to correctly identify an elusive reaction mechanism for the acid-catalyzed Morita–Baylis–Hillman reaction in methanol using this technique.[117] Additionally, a complete analysis was also performed for 11 different ionic species' solvation free energies ranging from 2+ to 2− charges that agreed with experimental values.[109]

In summary, we reaffirm that the intent of this Perspective is to stimulate development, benchmarking, and application of models used for studying mixed solvents. Over the years, there have been several challenges of predicting solvation free energies for drug and pesticide molecules in water (e.g., SAMPL), and in those, the COSMO-RS model has performed exceedingly well.[118,119] Some computationally intensive RISM implementations have also performed comparably well,[120] while cluster-continuum modeling has largely not been used except for single ion and small molecule solvation free energies. However, as explained above, modeling intermediates in chemical reactions (especially charged intermediates in electrochemical reactions) requires computational treatments with one of the above procedures. While COSMO-RS and RISM have shown strong performances in pure solvents, it is our perspective that cluster-continuum modeling with QCT and facilitated with the SOAP/sketch-map would be a useful treatment for studies in mixed solvents if and when COSMO-RS and RISM fail. The dual-QCT approach discussed above should make predicting solvation free energies of challenging systems tractable in mixed solvents without explicit simulations.

V. CONCLUSION

In this Perspective, we demonstrated that mixed solvents exhibit several theoretical and computational complexities that require attention. We discussed two known, but computationally challenging, characteristics of mixed solvents: the phase potential and preferential solvation. Phase potentials could impact computational studies with certain solvation treatments (e.g., cluster-continuum modeling) or inconsistent use of absolute or real solvation free energies. However, we primarily focused on current solvation models' potential to capture preferential solvation. Explicit solvation methods should produce the correct solvation free energies in mixed solvents provided sufficient sampling; however, this confidence comes at a significant computational cost of molecular dynamics with quantum chemistry or high quality force fields. Computationally cheaper hybrid methods (e.g., COSMO-RS and RISM) can more efficiently treat mixed solvent environments, but these methods have some known pitfalls, particularly with charged intermediates, that may limit their widespread use.

It is our perspective that additional solvation techniques should require limited *a priori* information without the need for *ab initio* molecular dynamics or well-parameterized force fields.

Cluster-continuum modeling, when rooted in QCT, has been useful for modeling charged intermediates, but the conventional QCT approach also requires explicit solvation to properly evaluate solvation free energies. We show here that the dual-QCT approach (leveraging global optimization techniques and SOAP/sketch-map analyses) should provide a suitable framework to account for preferential solvation in mixed solvents.

ACKNOWLEDGMENTS

REFERENCES

[1]C. Reichardt and T. Welton, *Solvents and Solvent Effects in Organic Chemistry* (John Wiley & Sons, 2011).

[2]B. M. Sato, C. G. de Oliveira, C. T. Martins, and O. A. El Seoud, "Thermo-solvatochromism in binary mixtures of water and ionic liquids: On the relative importance of solvophobic interactions," Phys. Chem. Chem. Phys. **12**, 1764–1771 (2010).

[3]K. Kwak, S. Park, and M. Fayer, "Dynamics around solutes and solute–solvent complexes in mixed solvents," Proc. Natl. Acad. Sci. U. S. A. **104**, 14221–14226 (2007).

[4]A. Farajtabar, F. Jaberi, and F. Gharib, "Preferential solvation and solvation shell composition of free base and protonated 5, 10, 15, 20-tetrakis(4-sulfonatophenyl)porphyrin in aqueous organic mixed solvents," Spectrochim. Acta, Part A **83**, 213–220 (2011).

[5]P. J. Dyson and P. G. Jessop, "Solvent effects in catalysis: Rational improvements of catalysts via manipulation of solvent interactions," Catal. Sci. Technol. **6**, 3302–3316 (2016).

[6]N. El-Mallah, S. Senior, G. Nabil, M. S. Ramadan, and E. Hamed, "Effect of acetonitrile–water mixtures on the reaction of dinitrochlorobenzene and dinitrochlorobenzotrifluoride with hydroxide ion," Int. J. Chem. Kinet. **42**, 453–463 (2010).

[7]V. Amenta, J. L. Cook, C. A. Hunter, C. M. Low, and J. G. Vinter, "Molecular recognition probes of solvation thermodynamics in solvent mixtures," Org. Biomol. Chem. **9**, 7571–7578 (2011).

[8]M. A. Mellmer, C. Sanpitakseree, B. Demir, P. Bai, K. Ma, M. Neurock, and J. A. Dumesic, "Solvent-enabled control of reactivity for liquid-phase reactions of biomass-derived compounds," Nat. Catal. **1**, 199–207 (2018).

[9]J. Ho and M. L. Coote, "A universal approach for continuum solvent pK_a calculations: Are we there yet?," Theor. Chem. Acc. **125**, 3–21 (2010).

[10]M.-H. Baik and R. A. Friesner, "Computing redox potentials in solution: Density functional theory as a tool for rational design of redox agents," J. Phys. Chem. A **106**, 7407–7412 (2002).

[11]C. Creutz, M. H. Chou, H. Hou, and J. T. Muckerman, "Hydride ion transfer from ruthenium(II) complexes in water: Kinetics and mechanism," Inorg. Chem. **49**, 9809–9822 (2010).

[12]A. Ben-Naim and Y. Marcus, "Solvation thermodynamics of nonionic solutes," J. Chem. Phys. **81**, 2016–2027 (1984).

[13]T. Lazaridis, "Inhomogeneous fluid approach to solvation thermodynamics. 1. Theory," J. Phys. Chem. B **102**, 3531–3541 (1998).

[14]D. Ben-Amotz, F. O. Raineri, and G. Stell, "Solvation thermodynamics: Theory and applications," J. Phys. Chem. B **109**, 6866–6878 (2005).

[15]M. Stöhr and A. Tkatchenko, "Quantum mechanics of proteins in explicit water: The role of plasmon-like solute-solvent interactions," Sci. Adv. **5**, eaax0024 (2019).

[16]M. Soniat, D. M. Rogers, and S. B. Rempe, "Dispersion- and exchange-corrected density functional theory for sodium ion hydration," J. Chem. Theory Comput. **11**, 2958–2967 (2015).

[17]A. V. Marenich, J. Ho, M. L. Coote, C. J. Cramer, and D. G. Truhlar, "Computational electrochemistry: Prediction of liquid-phase reduction potentials," Phys. Chem. Chem. Phys. **16**, 15068–15106 (2014).

[18]A. Marjolin and J. A. Keith, "Thermodynamic descriptors for molecules that catalyze efficient CO_2 electroreductions," ACS Catal. **5**, 1123–1130 (2015).

[19]A. A. Isse and A. Gennaro, "Absolute potential of the standard hydrogen electrode and the problem of interconversion of potentials in different solvents," J. Phys. Chem. B **114**, 7894–7899 (2010).

[20]C. P. Kelly, C. J. Cramer, and D. G. Truhlar, "Single-ion solvation free energies and the normal hydrogen electrode potential in methanol, acetonitrile, and dimethyl sulfoxide," J. Phys. Chem. B **111**, 408–422 (2007).

[21]J. Ho, M. L. Coote, C. J. Cramer, and D. G. Truhlar, "Theoretical calculation of reduction potentials," in *Organic Electrochemistry* (CRC Press, 2016), Chap. 6, pp. 229–259.

[22]S. Trasatti, "The absolute electrode potential: An explanatory note (Recommendations 1986)," Pure Appl. Chem. **58**, 955–966 (1986).

[23]M. D. Tissandier, K. A. Cowen, W. Y. Feng, E. Gundlach, M. H. Cohen, A. D. Earhart, J. V. Coe, and T. R. Tuttle, "The proton's absolute aqueous enthalpy and Gibbs free energy of solvation from cluster-ion solvation data," J. Phys. Chem. A **102**, 7787–7794 (1998).

[24]J. K. Nørskov, J. Rossmeisl, A. Logadottir, L. Lindqvist, J. R. Kitchin, T. Bligaard, and H. Jonsson, "Origin of the overpotential for oxygen reduction at a fuel-cell cathode," J. Phys. Chem. B **108**, 17886–17892 (2004).

[25]K. Leung, S. B. Rempe, and O. A. von Lilienfeld, "*Ab initio* molecular dynamics calculations of ion hydration free energies," J. Chem. Phys. **130**, 204507 (2009).

[26]S. B. Rempe and K. Leung, "Response to "Comment on '*Ab initio* molecular dynamics calculation of ion hydration free energies' [J. Chem. Phys. **133**, 047103 (2010)]"," J. Chem. Phys. **133**, 047104 (2010).

[27]M. I. Chaudhari, M. Vanegas, L. R. Pratt, A. Muralidharan, and S. B. Rempe, "Hydration mimicry by membrane ion channels," Annu. Rev. Phys. Chem. (in press) (2020).

[28]M. A. Wilson, A. Pohorille, and L. R. Pratt, "Surface potential of the water liquid–vapor interface," J. Chem. Phys. **88**, 3281–3285 (1988).

[29]T. L. Beck, "The influence of water interfacial potentials on ion hydration in bulk water and near interfaces," Chem. Phys. Lett. **561**, 1–13 (2013).

[30]E. Harder and B. Roux, "On the origin of the electrostatic potential difference at a liquid-vacuum interface," J. Chem. Phys. **129**, 234706 (2008).

[31]V. P. Sokhan and D. J. Tildesley, "The free surface of water: Molecular orientation, surface potential and nonlinear susceptibility," Mol. Phys. **92**, 625–640 (1997).

[32]J. Ho, "Predicting pK_a in implicit solvents: Current status and future directions," Aust. J. Chem. **67**, 1441–1460 (2014).

[33]S. M. Kathmann, I.-F. W. Kuo, C. J. Mundy, and G. K. Schenter, "Understanding the surface potential of water," J. Phys. Chem. B **115**, 4369–4377 (2011).

[34]M. D. Baer, A. C. Stern, Y. Levin, D. J. Tobias, and C. J. Mundy, "Electrochemical surface potential due to classical point charge models drives anion adsorption to the air–water interface," J. Phys. Chem. Lett. **3**, 1565–1570 (2012).

[35]T. S. Hofer and P. H. Hünenberger, "Absolute proton hydration free energy, surface potential of water, and redox potential of the hydrogen electrode from first principles: QM/MM MD free-energy simulations of sodium and potassium hydration," J. Chem. Phys. **148**, 222814 (2018).

[36]P. Debye and E. Hückel, "The theory of electrolytes. I. Lowering of freezing point and related phenomena," Phys. Z. **24**, 185–206 (1923).

[37] C. Hille, S. Ringe, M. Deimel, C. Kunkel, W. E. Acree, K. Reuter, and H. Oberhofer, "Generalized molecular solvation in non-aqueous solutions by a single parameter implicit solvation scheme," J. Chem. Phys. **150**, 041710 (2019).

[38] A. Ben-Naim, "Theory of preferential solvation of nonelectrolytes," Cell Biophys. **12**, 255–269 (1988).

[39] A. Ben-Naim, "Preferential solvation in two-component systems," J. Phys. Chem. **93**, 3809–3813 (1989).

[40] A. Ben-Naim, "Inversion of the Kirkwood–Buff theory of solutions: Application to the water–ethanol system," J. Chem. Phys. **67**, 4884–4890 (1977).

[41] Y. Marcus and Y. Migron, "Polarity, hydrogen bonding, and structure of mixtures of water and cyanomethane," J. Phys. Chem. **95**, 400–406 (1991).

[42] S. Weerasinghe and P. E. Smith, "A Kirkwood–Buff derived force field for methanol and aqueous methanol solutions," J. Phys. Chem. B **109**, 15080–15086 (2005).

[43] P. Ganguly and N. F. van der Vegt, "Convergence of sampling Kirkwood–Buff integrals of aqueous solutions with molecular dynamics simulations," J. Chem. Theory Comput. **9**, 1347–1355 (2013).

[44] Y. Marcus, "A quasi-lattice quasi-chemical theory of preferential solvation of ions in mixed solvents," Aust. J. Chem. **36**, 1719–1731 (1983).

[45] Y. Marcus, "Preferential solvation of ions in mixed solvents. Part 4—Comparison of the Kirkwood–Buff and quasi-lattice quasi-chemical approaches," J. Chem. Soc., Faraday Trans. 1 **85**, 3019–3032 (1989).

[46] Y. Marcus, *Ions in Solution and Their Solvation* (John Wiley & Sons, Inc., 2015).

[47] A. Ben-Naim, "Local composition and preferential solvation," in *Molecular Theory of Solutions* (Oxford University Press, 2006), Chap. 8, pp. 262–294.

[48] B. Cox, A. Parker, and W. Waghorne, "Coordination and ionic solvation," J. Phys. Chem. **78**, 1731–1740 (1974).

[49] A. K. Covington and K. E. Newman, "Approaches to the problems of solvation in pure solvents and preferential solvation in mixed solvents," Pure Appl. Chem. **51**, 2041–2058 (1979).

[50] P. Suppan, "Local polarity of solvent mixtures in the field of electronically excited molecules and exciplexes," J. Chem. Soc., Faraday Trans. 1 **83**, 495–509 (1987).

[51] A. Ben-Naim, *Molecular Theory of Solutions* (Oxford University Press, 2006).

[52] Y. Marcus, "Gibbs energies of transfer of anions from water to mixed aqueous organic solvents," Chem. Rev. **107**, 3880–3897 (2007).

[53] C. Kalidas, G. Hefter, and Y. Marcus, "Gibbs energies of transfer of cations from water to mixed aqueous organic solvents," Chem. Rev. **100**, 819–852 (2000).

[54] V. S. Bernales, A. V. Marenich, R. Contreras, C. J. Cramer, and D. G. Truhlar, "Quantum mechanical continuum solvation models for ionic liquids," J. Phys. Chem. B **116**, 9122–9129 (2012).

[55] A. Klamt, "Conductor-like screening model for real solvents: A new approach to the quantitative calculation of solvation phenomena," J. Phys. Chem. **99**, 2224–2235 (1995).

[56] A. Klamt, F. Eckert, and W. Arlt, "COSMO-RS: An alternative to simulation for calculating thermodynamic properties of liquid mixtures," Annu. Rev. Chem. Biomol. Eng. **1**, 101–122 (2010).

[57] M. Diedenhofen and A. Klamt, "COSMO-RS as a tool for property prediction of IL mixtures—A review," Fluid Phase Equilib. **294**, 31–38 (2010).

[58] A. Klamt, "The COSMO and COSMO-RS solvation models," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **1**, 699–709 (2011).

[59] A. Klamt, "The COSMO and COSMO-RS solvation models," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **8**, e1338 (2018).

[60] S. Sinnecker, A. Rajendran, A. Klamt, M. Diedenhofen, and F. Neese, "Calculation of solvent shifts on electronic g-tensors with the conductor-like screening model (COSMO) and its self-consistent generalization to real solvents (direct COSMO-RS)," J. Phys. Chem. A **110**, 2235–2245 (2006).

[61] C. C. Pye and T. Ziegler, "An implementation of the conductor-like screening model of solvation within the Amsterdam density functional package," Theor. Chem. Acc. **101**, 396–408 (1999).

[62] A. Klamt, "COSMO-RS for aqueous solvation and interfaces," Fluid Phase Equilib. **407**, 152–158 (2016).

[63] M. Diedenhofen, F. Eckert, and A. Klamt, "Prediction of infinite dilution activity coefficients of organic compounds in ionic liquids using COSMO-RS," J. Chem. Eng. Data **48**, 475–479 (2003).

[64] T. Gerlach, S. Müller, and I. Smirnova, "Development of a COSMO-RS based model for the calculation of phase equilibria in electrolyte systems," AIChE J. **64**, 272–285 (2018).

[65] A. Klamt, "COSMO-RSC: Second-order quasi-chemical theory recovering local surface correlation effects," J. Phys. Chem. A **120**, 2049–2056 (2016).

[66] D. Chandler and H. C. Andersen, "Optimized cluster expansions for classical fluids. II. Theory of molecular liquids," J. Chem. Phys. **57**, 1930–1937 (1972).

[67] D. Beglov and B. Roux, "An integral equation to describe the solvation of polar molecules in liquid water," J. Phys. Chem. B **101**, 7821–7826 (1997).

[68] E. L. Ratkova, D. S. Palmer, and M. V. Fedorov, "Solvation thermodynamics of organic molecules by the molecular integral equation theory: Approaching chemical accuracy," Chem. Rev. **115**, 6312–6356 (2015).

[69] A. Kovalenko, "Molecular theory of solvation: Methodology summary and illustrations," Condens. Matter Phys. **18**, 32601 (2015).

[70] R. Skyner, J. McDonagh, C. Groom, T. Van Mourik, and J. Mitchell, "A review of methods for the calculation of solution free energies and the modelling of systems in solution," Phys. Chem. Chem. Phys. **17**, 6174–6191 (2015).

[71] A. Kovalenko, "Multiscale modeling of solvation," in *Springer Handbook of Electrochemical Energy* (Springer-Verlag Berlin Heidelberg, 2017), Chap. 5, pp. 95–139.

[72] N. P. Reynolds, J. Adamcik, J. T. Berryman, S. Handschin, A. A. H. Zanjani, W. Li, K. Liu, A. Zhang, and R. Mezzenga, "Competition between crystal and fibril formation in molecular mutations of amyloidogenic peptides," Nat. Commun. **8**, 1338 (2017).

[73] J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids* (Elsevier, 1990).

[74] A. Kovalenko and F. Hirata, "A molecular theory of liquid interfaces," Phys. Chem. Chem. Phys. **7**, 1785–1793 (2005).

[75] 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).

[76] D. 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).

[77] Y. Kiyota, N. Yoshida, and F. Hirata, "A new approach for investigating the molecular recognition of protein: Toward structure-based drug design based on the 3D-RISM theory," J. Chem. Theory Comput. **7**, 3803–3815 (2011).

[78] S. Nishihara and M. Otani, "Hybrid solvation models for bulk, interface, and membrane: Reference interaction site methods coupled with density functional theory," Phys. Rev. B **96**, 115429 (2017).

[79] S. Gusarov, T. Ziegler, and A. Kovalenko, "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 **110**, 6083–6090 (2006).

[80] Z. Liu, C. Patel, J. N. Harvey, and R. B. Sunoj, "Mechanism and reactivity in the Morita–Baylis–Hillman reaction: The challenge of accurate computations," Phys. Chem. Chem. Phys. **19**, 30647–30657 (2017).

[81] L. R. Pratt and S. B. Rempe, "Quasi-chemical theory and implicit solvent models for simulations," AIP Conf. Proc. **492**, 172–201 (1999).

[82] D. Asthagiri, P. Dixit, S. Merchant, M. Paulaitis, L. Pratt, S. B. Rempe, and S. Varma, "Ion selectivity from local configurations of ligands in solutions and ion channels," Chem. Phys. Lett. **485**, 1–7 (2010).

[83] D. M. Rogers and S. B. Rempe, "Probing the thermodynamics of competitive ion binding using minimum energy structures," J. Phys. Chem. B **115**, 9116–9129 (2011).

[84] D. M. Rogers, D. Jiao, L. R. Pratt, and S. B. Rempe, "Structural models and molecular thermodynamics of hydration of ions and small molecules," in *Annual Reports in Computational Chemistry* (Elsevier, 2012), Vol. 8, pp. 71–127.

[85] T. L. Beck, M. E. Paulaitis, and L. R. Pratt, *The Potential Distribution Theorem and Models of Molecular Solutions* (Cambridge University Press, 2006).

[86] M. I. Chaudhari, L. R. Pratt, and S. B. Rempe, "Utility of chemical computations in predicting solution free energies of metal ions," Mol. Simul. **44**, 110–116 (2018).

[87] S. B. Rempe, L. R. Pratt, G. Hummer, J. D. Kress, R. L. Martin, and A. Redondo, "The hydration number of Li$^+$ in liquid water," J. Am. Chem. Soc. **122**, 966–967 (2000).

[88] S. B. Rempe and L. R. Pratt, "The hydration number of Na$^+$ in liquid water," Fluid Phase Equilib. **183–184**, 121–132 (2001).

[89] D. Asthagiri and L. R. Pratt, "Quasi-chemical study of Be^{2+}(aq) speciation," Chem. Phys. Lett. **371**, 613–619 (2003).

[90] D. Asthagiri, L. R. Pratt, M. E. Paulaitis, and S. B. Rempe, "Hydration structure and free energy of biomolecularly specific aqueous dications, including Zn^{2+} and first transition row metals," J. Am. Chem. Soc. **126**, 1285–1289 (2004).

[91] S. B. Rempe, D. Asthagiri, and L. R. Pratt, "Inner shell definition and absolute hydration free energy of K$^+$(aq) on the basis of quasi-chemical theory and *ab initio* molecular dynamics," Phys. Chem. Chem. Phys. **6**, 1966–1969 (2004).

[92] D. Sabo, D. Jiao, S. Varma, L. Pratt, and S. Rempe, "Case study of Rb$^+$(aq), quasi-chemical theory of ion hydration, and the no split occupancies rule," Annu. Rep. Prog. Chem., Sect. C: Phys. Chem. **109**, 266–278 (2013).

[93] M. I. Chaudhari, M. Soniat, and S. B. Rempe, "Octa-coordination and the aqueous Ba^{2+} ion," J. Phys. Chem. B **119**, 8746–8753 (2015).

[94] M. J. Stevens and S. L. B. Rempe, "Ion-specific effects in carboxylate binding sites," J. Phys. Chem. B **120**, 12519–12530 (2016).

[95] M. I. Chaudhari and S. B. Rempe, "Strontium and barium in aqueous solution and a potassium channel binding site," J. Chem. Phys. **148**, 222831 (2018).

[96] M. I. Chaudhari, S. B. Rempe, and L. R. Pratt, "Quasi-chemical theory of F$^-$(aq): The "no split occupancies rule" revisited," J. Chem. Phys. **147**, 161728 (2017).

[97] A. Muralidharan, L. R. Pratt, M. I. Chaudhari, and S. B. Rempe, "Quasi-chemical theory with cluster sampling from *ab initio* molecular dynamics: Fluoride (F$^-$) anion hydration," J. Phys. Chem. A **122**, 9806–9812 (2018).

[98] A. Muralidharan, L. R. Pratt, M. I. Chaudhari, and S. B. Rempe, "Quasi-chemical theory for anion hydration and specific ion effects: Cl$^-$(aq) *vs.* F$^-$(aq)," Chem. Phys. Lett.: X **4**, 100037 (2019).

[99] S. Mazur, "Neighborship partition of the radial distribution function for simple liquids," J. Chem. Phys. **97**, 9276–9282 (1992).

[100] S. Varma and S. B. Rempe, "Tuning ion coordination architectures to enable selective partitioning," Biophys. J. **93**, 1093–1099 (2007).

[101] D. Jiao and S. B. Rempe, "CO$_2$ solvation free energy using quasi-chemical theory," J. Chem. Phys. **134**, 224506 (2011).

[102] D. Sabo, S. Varma, M. G. Martin, and S. B. Rempe, "Studies of the thermodynamic properties of hydrogen gas in bulk water," J. Phys. Chem. B **112**, 867–876 (2008).

[103] S. Varma, D. Sabo, and S. B. Rempe, "K$^+$/Na$^+$ selectivity in K channels and valinomycin: Over-coordination versus cavity-size constraints," J. Mol. Biol. **376**, 13–22 (2008).

[104] S. Varma, D. M. Rogers, L. R. Pratt, and S. B. Rempe, "Design principles for K$^+$ selectivity in membrane transport," J. Gen. Physiol. **137**, 479–488 (2011).

[105] M. Rossi, A. Tkatchenko, S. B. Rempe, and S. Varma, "Role of methyl-induced polarization in ion binding," Proc. Natl. Acad. Sci. U. S. A. **110**, 12978–12983 (2013).

[106] M. I. Chaudhari, J. R. Nair, L. R. Pratt, F. A. Soto, P. B. Balbuena, and S. B. Rempe, "Scaling atomic partial charges of carbonate solvents for lithium ion solvation and diffusion," J. Chem. Theory Comput. **12**, 5709–5718 (2016).

[107] D. Jiao and S. B. Rempe, "Combined density functional theory (DFT) and continuum calculations of pK_a in carbonic anhydrase," Biochemistry **51**, 5979–5989 (2012).

[108] D. Jiao, K. Leung, S. B. Rempe, and T. M. Nenoff, "First principles calculations of atomic nickel redox potentials and dimerization free energies: A study of metal nanoparticle growth," J. Chem. Theory Comput. **7**, 485–495 (2011).

[109] Y. Basdogan, M. C. Groenenboom, E. Henderson, S. De, S. B. Rempe, and J. A. Keith, "Machine learning-guided approach for studying solvation environments," J. Chem. Theory Comput. **16**, 633–642 (2020).

[110] V. S. Bryantsev, M. S. Diallo, and W. A. Goddard III, "Calculation of solvation free energies of charged solutes using mixed cluster/continuum models," J. Phys. Chem. B **112**, 9709–9719 (2008).

[111] J. Zhang and M. Dolg, "ABCluster: The artificial bee colony algorithm for cluster global optimization," Phys. Chem. Chem. Phys. **17**, 24173–24181 (2015).

[112] J. Zhang and M. Dolg, "Global optimization of clusters of rigid molecules using the artificial bee colony algorithm," Phys. Chem. Chem. Phys. **18**, 3003–3010 (2016).

[113] A. P. Bartók, R. Kondor, and G. Csányi, "On representing chemical environments," Phys. Rev. B **87**, 184115 (2013).

[114] S. De, A. P. Bartók, G. Csányi, and M. Ceriotti, "Comparing molecules and solids across structural and alchemical space," Phys. Chem. Chem. Phys. **18**, 13754–13769 (2016).

[115] M. Ceriotti, G. A. Tribello, and M. Parrinello, "Simplifying the representation of complex free-energy landscapes using sketch-map," Proc. Natl. Acad. Sci. U. S. A. **108**, 13023–13028 (2011).

[116] M. Ceriotti, G. A. Tribello, and M. Parrinello, "Demonstrating the transferability and the descriptive power of sketch-map," J. Chem. Theory Comput. **9**, 1521–1532 (2013).

[117] Y. Basdogan and J. A. Keith, "A paramedic treatment for modeling explicitly solvated chemical reaction mechanisms," Chem. Sci. **9**, 5341–5346 (2018).

[118] J. Reinisch, A. Klamt, and M. Diedenhofen, "Prediction of free energies of hydration with COSMO-RS on the SAMPL3 data set," J. Comput.-Aided Mol. Des. **26**, 669–673 (2012).

[119] J. Reinisch and A. Klamt, "Prediction of free energies of hydration with COSMO-RS on the SAMPL4 data set," J. Comput.-Aided Mol. Des. **28**, 169–173 (2014).

[120] S. M. Kast, J. Heil, S. Güssregen, and K. F. Schmidt, "Prediction of tautomer ratios by embedded-cluster integral equation theory," J. Comput.-Aided Mol. Des. **24**, 343–353 (2010).

[121] B. Mennucci, E. Cancès, and J. Tomasi, "Evaluation of solvent effects in isotropic and anisotropic dielectrics and in ionic solutions with a unified integral equation method: Theoretical bases, computational implementation, and numerical applications," J. Phys. Chem. B **101**, 10506–10517 (1997).