Solvation Thermodynamics of Organic Molecules by the Molecular Integral Equation Theory: Approaching Chemical Accuracy

Ekaterina L. Ratkova,[†,‡] David S. Palmer,[¶,‡] and Maxim V. Fedorov[*,§,‡]

G.A. Krestov Institute of Solution Chemistry of the Russian Academy of Sciences, Akademicheskaya St. 1, 153045 Ivanovo, Russia, The Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, Leipzig, 04103, Germany, Department of Chemistry, University of Strathclyde, Thomas Graham Building, 295 Cathedral Street, Glasgow, Scotland G1 1XL, U.K., and Department of Physics, Scottish Universities Physics Alliance (SUPA), University of Strathclyde, John Anderson Building, 107 Rottenrow East, Glasgow G4 0NG, United Kingdom

E-mail: maxim.fedorov@strath.ac.uk

Phone: +44 141 54 84 012. Fax: +44 141 55 22 891

Contents

[*]To whom correspondence should be addressed
[†]G.A. Krestov Institute of Solution Chemistry of the Russian Academy of Sciences
[‡]Max Planck Institute for Mathematics in the Sciences
[¶]University of Strathclyde, Chemistry
[§]University of Strathclyde, Physics

1 Introduction

The Integral Equation Theory (IET) of Molecular Liquids has been an active area of academic research in theoretical and computational physical chemistry for over 40 years because it provides a *consistent* theoretical framework for describing the structural and thermodynamic properties of liquid-phase solutions. The theory can describe pure and mixed solvent systems (including *anisotropic* and *non-equilibrium* systems) and has been already used for theoretical studies of a vast range of problems in chemical physics/physical chemistry, molecular biology, colloids, soft matter and electrochemistry. A considerable advantage of IET is that it can be used to study specific solute-solvent interactions, unlike continuum solvent models, but yet it requires considerably less computational expense than explicit solvent simulations.

However, until recently this area of research (although active) was mostly considered as an outlier compared to molecular simulation methods. Among other reasons for this, we would like to highlight the following: (i) due to several problems with bridge functionals (see below), the theory has traditionally been considered to be too inaccurate for widespread use in practical applications such as research in the biomedical and environmental sciences; (ii) a lack of stable implementations of the IET algorithms in user-friendly computational software prevented researchers with non-computational backgrounds from using these methods.

The situation has changed during the last decade. Recent developments in theoretical and computational aspects of IET have made it possible to make accurate calculations of thermodynamic and structural properties of solvation across multiple classes of molecular systems at relatively low computational expense.[1–9] IET methods have been implemented in several open-source and proprietary pieces of computational chemistry software including Amber,[1] ADF[10] and MOE.[11] Furthermore, IET has found an ever-increasing number of successful applications, including computing solubility of druglike molecules,[12] fragment-based drug design,[13,14] molecular recognition,[15] modelling the binding of water and ions by proteins and DNA,[16–20] predicting tautomer ratios,[21] interpreting solvent densities around biomacromolecules,[22] and sampling molecular conformations.[3,7,23–27] As a result, more and more groups around the world have become involved in either

using the existing IET methods or developing new ones, and it is anticipated that the popularity of IET methods will rapidly increase in the near future.

The main purpose of this review is to overview the recent developments in IET with a particular focus on IET methods for predicting solvation thermodynamics of organic molecules (including biomolecules), an area that has undergone significant developments in the last 5-7 years. Because the Reference Interaction Site Model (RISM) is (arguably) the most developed *molecular* IET method in terms of computational and practical aspects, we will mainly focus on it in the review. Other molecular theory methods (classical DFT, Molecular Ornstein Zernike and energy-representation method) will be briefly overviewed at the end of the manuscript.

We note that most of the studies we will discuss in the manuscript deal with aqueous solutions due to the importance of the subject. However, other liquid systems will also be considered in various sections of the review where they are relevant.

To set the scene, we will start with basic experimental and theoretical aspects of solvation thermodynamics followed by a brief description of the main formulae of the IET. Then we will discuss practical aspects of the RISM methods (numerical algorithms, software implementations and hybridisation of RISM with other methods). Thermodynamic calculations within RISM will be overviewed in Section 5 where we will focus on new solvation free energy functionals developed during the last few years. Different applications of RISM will be discussed in Section 6 followed by a section on other molecular theory methods. We finish our review with Conclusions where we will discuss ongoing trends, current challenges and prospective directions for future developments in IET.

2   Solvation thermodynamics: general concepts

Ben-Naim defined *solvation* as the process of transferring one solute molecule from fixed position in an ideal gas phase to a fixed position in the liquid phase at constant pressure and temperature.[28] In the case of aqueous solutions, the solvation process is called *hydration*.

Below we provide basic definitions of the main thermodynamics quantities that characterise processes associated with molecular solvation with a focus on solvation free energy (SFE) and its links with other important thermodynamics quantities. We will also briefly discuss different experimental techniques for measuring SFE in a view to assessing the validity of available experimental data in the literature. At the end of the section, we will briefly overview the main modern non-IET computational approaches to calculate SFE.

2.1 Solvation free energy (SFE) and its key role in solvation thermodynamics

The change of Gibbs free energy of a molecular solute upon solvation - *solvation free energy*, SFE, is a fundamental equilibrium quantity which is inherently related to many other important thermodynamic characteristics like partition coefficient, solubility, association/dissociation constant, etc. In the case of aqueous solutions, the SFE is usually called *hydration free energy* (HFE).

2.1.1 SFE and partition coefficients

In chemistry, a partition coefficient (P) is the ratio of the concentrations of a substance in two heterogenous phases in thermodynamic equilibrium with each other.[30] In the case of partition of a molecule between gaseous phase and aqueous solution the partition coefficient is called the Henry's law constant (K_H):[31,32]

$$K_H = \frac{C_{g(i)}}{C_{aq(i)}} \tag{1}$$

where $C_{g(i)}$ and $C_{aq(i)}$ are equilibrium molecular concentrations of the molecules i in gaseous and aqueous phases, respectively. (The inverse of this equilibrium constant is sometimes referred to as the Ostwald solubility coefficient.) The relationship between HFE and K_H is:

$$\Delta G_{hyd} = RT \ln (K_H), \tag{2}$$
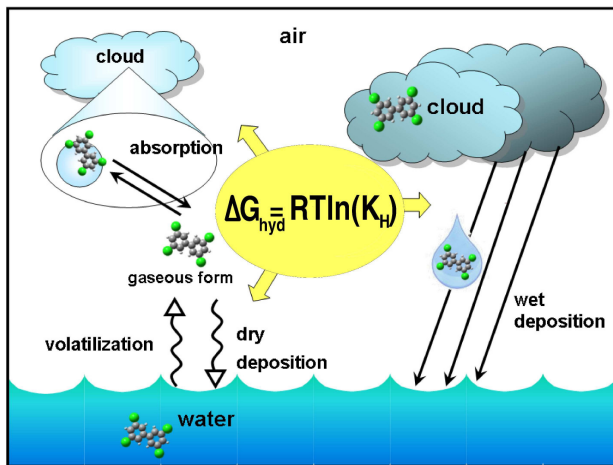
Figure 1: The main pathways that influence the distribution of persistent organic pollutants (POPs) between atmosphere and water. Hydration free energy (ΔG_{hyd}) is an important thermodynamic parameter used to describe the main processes of POP distribution between atmosphere and water. It is closely related to the Henry's law constant, K_H (see eq 2). In turn, K_H is widely used to model the flux of a molecule from air to water, $F_{1\to 2}$ (see eq 3). Reprinted with permission from Ref.[29] Copyright 2011 American Chemical Society.

where ΔG_{hyd} is the HFE, R is the ideal gas constant, T is the temperature. Experimental methods to obtain Henry's law constants are discussed below (see the Section Experimental methods to measure SFE). Henry's law constants are widely used in environmental chemistry to describe compound exchange between air (troposphere) and water (ocean, river, lake, etc.)[33–37] Several dominant mechanisms that determine the air-water exchange are shown in Figure 1. Generally, they can be modeled using the equation of a *flux across surfaces*:[34,37]

$$F_{1\to 2(i)} = K_{1/2(i)} \left(C_{1(i)} - \frac{C_{2(i)}}{P_{i,eq}} \right), \tag{3}$$

where $F_{1\to 2(i)}$ is the flux of molecules i from compartment 1 to compartment 2, $K_{1/2(i)}$ is the kinetic parameter represented by the mass transfer coefficient of the molecules i, $C_{1(i)}$ and $C_{2(i)}$ are equilibrium molecular concentrations of the molecules i in compartments 1 and 2, respectively, $P_{i,eq}$ is the equilibrium partition coefficient of the molecules i between the two compartments. The air-water exchange is particularly important to investigate the global fate of persistent organic pollutants (POPs).[34,38–40] This class of pollutants is characterized by: (i) long-term persistence,

7
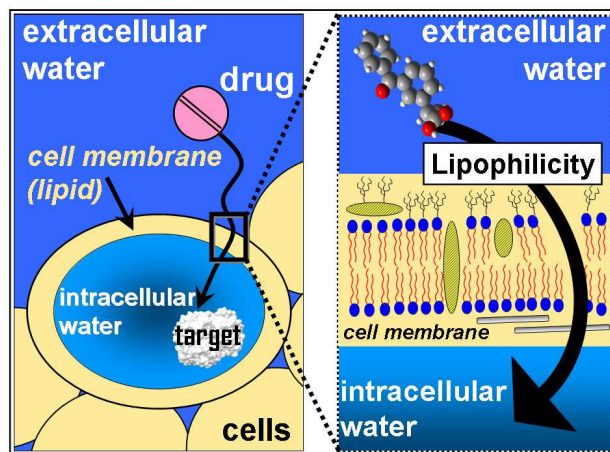
Figure 2: To reach the target protein drug molecules should overcome several physical-chemical barriers. One of them is penetration of the molecules, situated in the intercellular aqueous media, through lipid cell membrane. The property which shows the molecule partition behavior between these two media as called lipophilicity. It was found[42] that the value of logarithm of partition coefficient between n-octanol and water ($\log P_{oct/wat}$) has high correlation with the corresponding value of lipophilicity.

(ii) long-range atmospheric transport and deposition, (iii) bioaccumulation, (iv) adverse effects on biota.[34,38] POPs have been used as fungicides and pesticides for many years.[38] Although the further use and production of POPs has been banned,[41] their persistence in biological compartments (e.g. soil, water, plants, and sediment) means that they still pose a significant environmental hazard. The global extent of the distribution of POPs became apparent with their detection in the Arctic and Antarctic (where they have never been used or produced) at levels posing a hazard for both wildlife and humans.[38] For this reason, understanding and clarifying the global fate of POPs is one of the most important current environmental and ecological problems.

In turn, the logarithm of the partition coefficient between n-octanol and water ($\log P_{oct/wat}$) is a key physical/chemical parameter in understanding the distribution of molecules in both environmentally and pharmaceutically relevant systems. The value of $\log P_{oct/wat}$ is highly correlated with the value of lipophilicity[42] which is commonly defined as the partitioning behavior of compounds between aqueous and organic phases (fats, membrane lipids, etc.) (see Figure 2). Lipophilicity is an essential parameter to estimate absorption, transport, and excretion of pollutants and druglike

compounds from a living organism. Log $P_{oct/wat}$ is defined as the decadic logarithm of the ratio of the concentration of the un-ionized form of a compound in n-octanol to the concentration of its un-ionized form in aqueous phase:

$$\log P_{oct/wat} = \log \left(\frac{[A]_{oct}^{un-ionized}}{[A]_{wat}^{un-ionized}} \right), \tag{4}$$

where $[A]$ is the equilibrium concentration of compound A. To measure the partition coefficient of ionizable solutes, the pH of the aqueous phase is adjusted such that the predominant form of the compound is un-ionized. The classical and most reliable technique for measurement of $\log P_{oct/wat}$ is the shake-flask method, which consists of dissolving a fixed quantity of the solute in a volume of n-octanol and water, then measuring the concentration of the solute in each solvent. The method is well-described elsewhere.[43,44] Alternatively, $\log P_{oct/wat}$ can be estimated using values of the corresponding free energies of hydration and solvation in n-octanol:

$$\log P_{oct/wat} = -\frac{\Delta G_{solv(oct)} - \Delta G_{hyd}}{RT \ln(10)}, \tag{5}$$

where the solvation free energy in n-octanol is given as $\Delta G_{solv(oct)}$. Although $\log P_{oct/wat}$ in this equation does not correspond exactly to that measured by experiments, because the miscibility of water and n-octanol is not zero, it has been widely used and is often considered to be satisfactory to give useful first order estimates of n-octanol/water partition coefficients (or, conversely, of hydration or n-octanol solvation free energies).

2.1.2 SFE and association/dissociation constants

Molecular association processes (complex formation, self-assembling, etc.) are essential in many fields of investigation: chemistry, biochemistry, pharmaceutics, food engineering, etc. To give an example, excessively strong binding of a drug molecule by blood proteins may significantly decrease the concentration of free (non-bound) drug molecules in the blood stream resulting in lower efficiency of the pharmaceutical against a disease.[45] The strength of binding (association) is
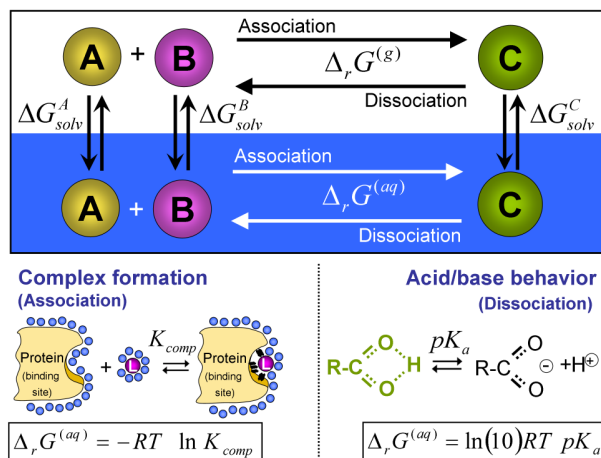
Figure 3: Thermodynamic cycle to estimate constants of association (e.g. protein-ligand binding) and dissociation (e.g. carboxylic group dissociation) in liquid (aqueous) phase $\Delta_r G^{(aq)}$ using the corresponding value in gaseous phase $\Delta_r G^{(g)}$ and solvation (hydration) free energies for reagents and products ΔG_{solv}^A, ΔG_{solv}^B, ΔG_{solv}^C.

commonly evaluated by the complex formation constant (K_{comp}) (also called the stability constant). We note that biochemists rarely talk about association constants, but rather their reciprocal, the dissociation constants (K_d). In the special case of salts, the dissociation constant can also be called an ionization constant (K_i). The classical experimental techniques to determine the constants (e.g. differential scanning calorimetry, isothermal titration calorimetry, etc) are well-described elsewhere.[46–48] The value of K_{comp} can be calculated indirectly from the thermodynamic cycle shown in Figure 3, in which the hydration/solvation free energy terms are important terms. For example, for the process of protein-ligand binding, K_{comp} can be calculated as:

$$\ln K_{comp} = -\frac{\Delta_r G^{(g)} + \Delta G_{hyd}^{PL} - \left(\Delta G_{hyd}^P + \Delta G_{hyd}^L\right)}{RT}, \qquad (6)$$

where K_{comp} is the complex formation constant, $\Delta_r G^{(g)}$ is the free energy of association (e.g. protein - ligand binding) in gaseous phase, ΔG_{hyd}^P, ΔG_{hyd}^L, and ΔG_{hyd}^{PL} are hydration free energies of protein (P), ligand (L), and protein-ligand complex (PL), respectively.

The thermodynamic cycle via the vapour (see Figure 3) may also be used to study acid-base behavior of pharmaceutical molecules, which is of key importance in estimating their distribution in an organism. Many physical/chemical properties are extremely sensitive to whether the drug

10

molecule is neutral or charged: aqueous solubility, lipophilicity, membrane permeability - to name a few. The acidity of molecules is usually estimated by the acid dissociation constant (K_a). Due to the many orders of magnitude spanned by K_a values, a logarithmic measure of the acid dissociation constant is more commonly used in practice: $pK_a = -\log_{10} K_a$. The classical experimental techniques of direct pK_a determination (potentiometric titration, spectrophotometric method, conductometry, etc.) are well-described elsewhere.[49,50] Alternatively, pK_a can be calculated using solvation free energy terms (see Figure3). For the case of pK_a in water:

$$pK_a = \frac{\Delta_r G^{(g)} + \Delta G_{hyd}^{A^-} + \Delta G_{hyd}^{H^+} - \Delta G_{hyd}^{HA}}{RT \ln(10)}, \tag{7}$$

where pK_a is the acid dissociation constant, $\Delta_r G^{(g)}$ is the free energy of the reaction (dissociation of the acid HA) in gaseous phase,ΔG_{hyd}^{HA}, $\Delta G_{hyd}^{A^-}$, and $\Delta G_{hyd}^{H^+}$ are HFEs of protonated acid HA, anion A^-, and proton H^+, respectively, R is the ideal gas constant, T is the temperature. We note that as the free energy of reaction in solution phase is a state function its value can be obtained from other thermodynamic cycles.[51]

2.1.3 SFE and intrinsic solubility

The intrinsic solubility of an ionizable molecule is defined as the concentration of the neutral form of the molecule in solvent at thermodynamic equilibrium with a small excess of undissolved solute. For a crystalline solute, solvation free energy can be related to intrinsic solubility by a thermodynamic cycle via the vapor:[52]

$$\Delta G_{sol} = \Delta G_{sub} + \Delta G_{solv} = -RT \ln (V_m S_o), \tag{8}$$

where ΔG_{sol} is the Gibbs free energy of solution, ΔG_{sub} is the Gibbs free energy of sublimation ΔG_{solv} is the SFE, S_o is the intrinsic solubility, R is the molar gas constant, T is the temperature, and V_m is the molar volume of the crystalline solute. The calculation of the intrinsic solubility of organic molecules via computation of the Gibbs free energy of solution remains a significant chal-

lenge for computational chemistry since an error of 5.7 kJ/mol (1.36 kcal/mol) in ΔG_{sol} equates to a 10-fold error in solubility. It has recently been shown that implicit continuum or explicit solvent simulation give random errors of $\sim$10-15 kJ/mol ($\sim$2.5-3.5 kcal/mol) when used to calculate the SFEs of *druglike* molecules.[53,54] Nevertheless, some success has been achieved using this approach for simple organic solutes[55] and for druglike molecules.[12,52]

2.1.4 Enthalpic and entropic terms

SFE contains two contributions:[56]

$$\Delta G_{solv} = \Delta H_{solv} - T\Delta S_{solv}, \tag{9}$$

where ΔH_{solv} is the change of enthalpy and ΔS_{solv} is the change of entropy for the solvation process. The first term (ΔH_{solv}), the so-called heat effect of the solvation process, equals the amount of energy that can be converted to heat (useful work) at constant pressure and temperature (it can be measured calorimetrically), whereas the last term ($T\Delta S_{solv}$) equals the amount of energy that the system gives to an external medium due to its disordering (or, otherwise, the system takes from the external medium for its ordering). So, the system should be in a thermal contact with the external media, the temperature of which is constant and equal to the temperature of the system. For example, let us consider a hydration complex formation upon solvation of a molecule with polar/charged groups or a metal ion in water.[57,58] The process results in a more ordered system due to the formation of additional bonds between the solute and water and the entropy of the system decreases ($\Delta S_{solv} < 0$, $-T\Delta S_{solv} > 0$).[56] It means that the system takes some energy from the external medium for molecule-water bond formation. However, complex formation is an exothermic process and is accompanied by the release of heat ($\Delta H < 0_{solv}$). Due to the fact that, generally, during the formation of a hydration complex the amount of heat released is larger than the amount of absorbed energy,[57,59] the free energy of the reaction has a negative value, $\Delta G_{solv} < 0$, indicating that the process is spontaneous.

2.1.5 Free energy profiles

Solvation and related processes (e.g. dissociation/association of complexes in solution) can often be characterized by a free energy profile (a so-called potential of mean force, PMF). According to the Ben-Naim definition, [28] PMF is the work involved (the Gibbs free energy at constant pressure and temperature) in bringing two (or many) selected particles from infinite separation (in condensed phase) to the final configuration (in condensed phase). One should note that the PMF is closely related to the pair density distribution function which gives the probability of finding a particle j at a certain distance from particle i:

$$PMF(\mathbf{r}_{ij}, \Omega_{ij}) = -k_B T \ln \left[g(\mathbf{r}_{ij}, \Omega_{ij}) \right], \tag{10}$$

where $g(\mathbf{r}_{ij}, \Omega_{ij})$ is the pair density distribution function, $\mathbf{r}_{ij}$ is the distance-vector between i^{th} and j^{th} molecules, $\Omega_{ij} = \{\phi_{ij}, \theta_{ij}, \psi_{ij}\}$ is the relative orientations of the ith and j^{th} molecules (given by the set of Euler angles), k_B is the Boltzmann constant, T is the temperature. For a complex formation process in water (e.g. protein – ligand binding), the free energy profile[60] and its enthalpic and entropic contributions are shown in Figure 4. At a distance A, there are no interactions between the protein and the ligand because the molecules are far away from each other. As the ligand comes closer to the protein binding center (Figure 4, distance B), both of them must release some water from their first solvation shells due to steric crowding. The process is accompanied by an increase of the system entropy ($\Delta_r S^{(aq)} > 0$, $-T\Delta_r S^{(aq)} < 0$) and the absorption of heat from external media for the partial desolvation of the interacting molecules ($\Delta_r H^{(aq)} > 0$). However, the distance B is still too large for the formation of direct contacts between the ligand and the protein, which takes place at a distance C. The heat release due to the protein - ligand interactions and the rearrangement of water molecules is considerable larger than the increase of the entropic term (entropy of the system decreases). So, the resulting free energy of the interactions is negative indicating that the process is spontaneous.
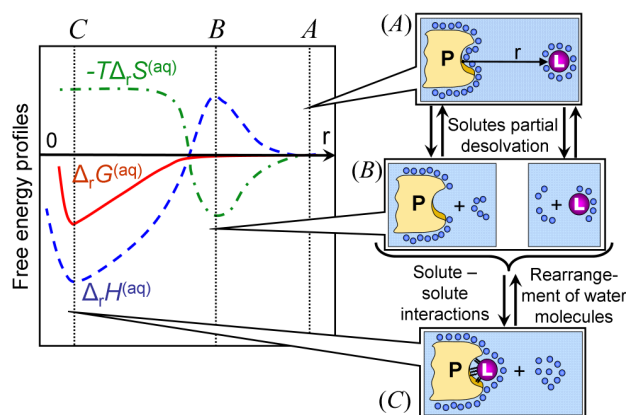
Figure 4: Free energy profile (red solid line) and its enthalpic (blue dashed line), and entropic (green dashed-dotted line) contributions of the protein – ligand binding in water, where **r** is the distance between the ligand (L) and the binding site of protein (P). Behavior of the system at distances A, B, and C is discussed in the text. The free energy profiles were adopted with permission from Ref.[60] Copyright 2010 American Chemical Society.

Conclusion: As the previously discussed physical/chemical properties are widely used in environmental chemistry to predict the global fate of persistent organic pollutants[34,39] and in pharmaceutical chemistry in predicting the pharmacokinetic behavior of novel drug-like molecules (e.g. tablet stability and dissolution, oral digestion, membrane penetration, and absorption in different tissues[52,61–64]), accurate and fast methods for determination of solvation/hydration free energies would have wide-spread benefits.

2.2   Experimental methods to measure SFE

The common experimental techniques to measure the solvation free energies of organic molecules can be roughly separated into *direct* and *indirect* methods.

The choice of one or another approach depends on the solubility and volatility of the compounds under investigation. Generally, the solubility of both pharmaceutically- and ecologically-relevant compounds in water is low and one may assume (or make a correction to) an infinitely-diluted solution. In this case, HFE can be estimated through the Henry's law constant (see eq 2). Direct methods are applicable only for molecules with volatility from relatively high $(C_{gas}/C_{aq} > 0.1)$ to moderate $(0.10 > C_{gas}/C_{aq} > 0.01)$,[65] e.g. persistent organic pollutants. The compounds with

14

lower volatility (e.g. drug-like molecules) can be investigated by indirect approaches that calculate solvation free energies from separate measurements of intrinsic solubility and pure compound vapour pressures. We will briefly review both direct and indirect methods for measurement of SFEs with a focus on aqueous solutions.

2.2.1 Direct methods

The direct methods can be categorised as either *static* or *dynamic*. Static methods involve preparation of dilute aqueous solutions containing an organic compound and allowing it to partition into the gas phase (so-called headspace) in contact with the solution in a closed vessel under temperature-controlled conditions until thermodynamic equilibrium is reached. The Henry's law constant determination within the approach is based on the following equation:

$$C_g = C_{aq_0} \frac{K_H(V_g/V_{aq})}{K_H(V_g/V_{aq}) + 1},$$ (11)

where C_g is the headspace concentration of an organic solute, C_{aq_0} is the initial concentration of the solute, K_H is the Henry's law constant, V_g and V_{aq} are the volumes of the headspace and aqueous phases, respectively.

In this approach, the error in determining K_H can be reduced by one of two widely used techniques: the equilibrium partition in closed system (EPICS)[66] approach or the static headspace method (the latter is also called phase ratio variation, PRV).[67] EPICS is based on measurement of the headspace concentration of a solute from multiple sealed bottles having different headspace-to-liquid volume ratios (V_g/V_{aq}) and the same initial amount of the solute (the initial number of mole is constant). The principle problem limiting the accuracy of the EPICS method is the difficulty of delivering equal quantities of compound in the several bottles. In turn, the static headspace method considers the equilibration in one vial of a given volume of an aqueous solution (C_{aq_0} =constant) with several given volumes of solute-free air (different amounts of the aqueous solution are removed from the vial with a syringe).

Generally, direct static techniques are sensitive enough for compounds with high volatility

because of the limited size of the headspace and water sample. For molecules with moderate volatility, the direct dynamic approach is preferable. The technique is more accurate than the static equilibration method because the measurement is based on a relative concentration change in one phase, and as a result, it is also capable of measuring much lower values of C_{gas}/C_{aq}. The direct dynamic approach include such techniques as the gas stripping method (GSM)[68] and the so-called "wetted-wall column" (WWC) or concurrent flow technique.[69]

The first of these methods allows the determination of the Henry's law constant by measurement of the compound loss in liquid phase by stripping it from the solution into a gaseous phase using a bubble column apparatus:

$$\ln\left(\frac{C_i}{C_0}\right) = -\frac{K_H G}{V R T} t_i, \tag{12}$$

where C_i and C_0 are current and initial concentrations of the solute, G is the gas flow rate, V is the volume of the liquid, t_i is the current time, R is the gas constant, T is the temperature. This experimental design requires that the velocity of the rising bubbles is sufficiently small and the height of the column is sufficiently great to establish air-water equilibrium. Moreover, the bubbles need to be large enough to allow the absorption of the solute molecules to the surface of the gas bubbles to be neglected.[70,71]

An alternative dynamic approach is the WWC technique in which the compound of interest is equilibrated between a flow of gas and a thin layer of solvent within a long vertical column.[72] This technique allows one to avoid the problem of solute absorption; however, there is a risk of trapping solution aerosol in the vapor trap at the bottom of the mixing column.[69] As with the gas stripping method, it is important that sufficient contact time is allowed to ensure that chemical equilibrium has been reached.

Thus, for compounds with an extremely low volatility (e.g. drug-like molecules) even the dynamic approach is not sensitive enough.

$$\Delta G_{solv} = \Delta G_{soln} - \Delta G_{vap/sub}$$
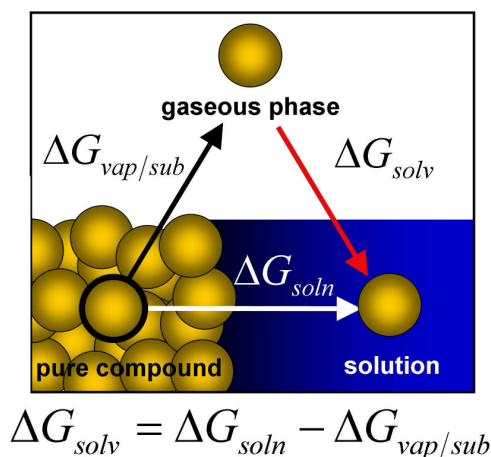
Figure 5: Thermodynamic cycle to estimate solution free energy (ΔG_{soln}) using the corresponding value of the vaporization free energy (ΔG_{vap}) or sublimation free energy (ΔG_{sub}) and solvation free energy (ΔG_{solv}). The figure was adopted with permission from Ref.[233] Copyright 2011.

2.2.2 Indirect methods

The solvation free energy of solutes with low vapour pressures can often be calculated indirectly from separate measurements of solubility and pure compound vapour pressure by a thermodynamic cycle via the vapour (see Figure 5).

The traditional "gold-standard" method for measuring solubility is the *shake-flask method*. The sample compound is added to a buffered solution until precipitation occurs. The solution is then agitated until thermodynamic equilibrium has been established. The excess precipitate is removed by filtration or centrifugation, and the concentration of compound in solution is measured. The main problem with the shake-flask method is that it is difficult to guarantee that thermodynamic equilibrium has been established. Additional errors may also be incurred by the filtration or centrifugation step and the use of buffers. Alternatively, solubility can be measured by *potentiometric methods* which are more accurate than many classical methods.[73–75] This makes them suitable for use in a pharmaceutical development setting, where accurate solubility data is required.[76,77] The principle benefit of the potentiometric methods is that they increase the rate at which thermodynamic equilibrium is obtained, but they can typically only be used for solutes with a pK_a in the range 2 - 12. Experimental intrinsic solubility data reported as part of a blind challenge to predict

solubilities was measured by a potentiometric method.[78]

Reliable experimental vapor pressure data for a wide range of structurally simple compounds (generally, with intermediate vapor pressure, 1 – 100 kPa) are available from established databases (e.g. Dortmund Databank, DDB) and published compilations.[79–81] However, these databases do not contain experimental data for compounds with ambient vapor pressures ¡∼ 0.0001 kPa (e.g. pharmaceutical molecules).[82] Although moderate vapour pressures (1 – 200 kPa) are relatively easy to measure, it remains a significant challenge to measure low (0.001 – 1 kPa) and very low (¡ 0.001 kPa) vapor pressures. The standard experimental techniques to measure vapor pressure in low-pressure regions are the transpiration method (or inert gas flow method)[83] and the molecular (Knudsen) effusion method. In *the transpiration method* a stream of inert gas is passed over the sample at a constant temperature and known flow rate such that the carrier gas becomes saturated with the sublimed sample. The carrier gas is then condensed downstream and the mass and purity of the sublimate are measured. The vapour pressure over the pure sample at that temperature is then calculated from the volume of the inert gas and the mass of the sublimate. The main problem faced in these experiments is that bioactive molecules typically have very low vapour pressures that are difficult to measure accurately at room temperature. Furthermore, one has to be cautious about the chemical stability of the solutes under investigation.[84] In *Knudsen effusion* a sample is placed in a closed container ("Knudsen cell") with a small hole (0.5 - 1.5 mm diameter) in the side. It is assumed that thermodynamic equilibrium is established between the sample and its vapor phase inside the container. The vapor effuses out of the small hole into vacuum, and the vapor pressure is calculated using the Knudsen equation by measuring the rate of the sample mass loss.

2.2.3 Available experimental SFE data

Currently, experimental SFE data are available for only several thousand of the millions of known organic compounds and around a hundred solvents.[85–92] The largest available compilations of SFEs for small molecules is the Minnesota Solvation Database ($http://comp.chem.umn.edu/mnsol/$), which presents a collection of 3037 experimental free energies of solvation or transfer free energies

for 790 unique solutes (541 neutral solutes and 249 singly-charged ions) in 92 solvents (including water). The distributed version of the database requires a licence, which is free of charge for non-profit organizations.

Chamberlin, Cramer and Truhlar collected a dataset of experimental HFEs at varied temperatures.[93,94] The dataset contains measurements of HFEs from different sources for a variety of small, non-ionic organic molecules in a broad temperature range and was used for parameterization of the SM6T and SM8T solvation models.[93,94]

Recently D.L. Mobley and J.P. Guthrie organized the FreeSolv database ($http://www.escholarship.org/uc/item/6sd403pz$), which builds on previous work from the Mobley lab and others. It contains experimental and calculated HFE for 504 small molecules. Experimental values are taken from prior literature and will continue to be curated. Calculated values are based on the GAFF small molecule force field in TIP3P water with AM1-BCC charges. We note that very few experimental SFE data are available for druglike molecules. The available collections of HFEs for drug-like and pesticide-like molecules are the SAMPL1,[54] SAMPL2,[53] SAMPL3[95] and SAMPL4[96] datasets recently published as part of blind prediction challenges.

Conclusion: Although there has been recent progress in the development of experimental methods to obtain SFEs, accurate measurements of SFEs for drug-like molecules and pollutants are still problematic. Experimental investigation of compounds with low volatility and/or solubility are challenging because they require precise and laborious measurements with highly sensitive instruments. As a result, these studies remain expensive and time-consuming (about one month of intensive work is necessary to determine the SFE of one drug-like molecule).[97]

2.3 SFE calculations: brief overview of main approaches

Accurate calculation of SFEs/HFEs of organic molecules has been a subject of very intensive research in computational and theoretical chemistry for many years. Indeed, as a fundamental thermodynamic parameter, SFE is a key quantity to verify the validity of a solvation theory or

method. From another side, SFE is important in many applied aspects of research in computational chemistry. For example, SFEs have been used in the calculation of octanol-water partition coefficients,[61,98,99] protein-ligand binding affinities,[100–102] acid-base dissociation constants (pKa's),[103] and aqueous solubilities.[12,52,104] Since these physicochemical properties are important in industrial applications, such as predicting the pharmacokinetic behavior of novel pharmaceutical molecules, improving the accuracy of computational methods to calculate SFEs would have widespread benefits.

Many different methods for SFE/HFE predictions have been developed at different levels of complexity: from empirical relationships like linear free energy and QSPR models and to 'bottom up' QM/MM methods with explicit solvent models. Overviewing all of them is out of scope of this review; we will only briefly overview the main classes of SFE computational methods in this section. For more information we refer to the following reviews:

- We would like to emphasise the role of recent SAMPL1,[54] SAMPL2,[53] SAMPL3,[95] and SAMPL4[96] blind HFE challenges that provoked a wave of interest in HFE predictions using different computational techniques. These works contain several benchmark data sets that can be used for verifying new models. The critical analyses of the performance of different methods in these challenges are insightful and provide interesting information.

- Critical comparisons of the HFE predictive ability for explicit and implicit solvent models can be found in Refs.[54,91,105–115]

2.3.1 'Top down' methods: cheminformatics and QSPR approaches

Quantitive Structure-Property Relationships (QSPRs) are empirically parameterized functions that relate a target property to a set of molecular descriptors calculable from a simple computational representation of the molecule.[116]

The simplest form of a QSPR model used to predict SFE are the Group Contribution (GC)

methods,[87,115] in which the SFE is commonly represented as a linear sum:

$$\Delta G_{solv} = \sum a_i x_i + c, \tag{13}$$

where a_i and c are the regression coefficients and intercept determined by multi-linear regression, x is a count of the number of occurrences of a particular atom or fragment in the solute and ΔG_{solv} is the SFE, which is the property to be predicted. The GC methods are a form of linear free energy relationship.[117–119] For small molecule datasets, GC methods often report quite accurate predictions, but they have a number of well-known problems, which mean that they often do not extrapolate well to larger and more complicated molecules. Firstly, atomic or molecular fragments are not necessarily independent of each other, their effect on intra- and inter- molecular interactions will depend upon their situation in the molecule and their environment. Secondly, the assumption of linearity is questionable, i.e. the addition of a second or third hydrophobic group to a molecule may have a different effect on its HFE than addition of a first hydrophobic group. There are also a number of practical problems with Group Contribution methods. In order to determine a statistically significant regression coefficient for each variable and to ensure that all common molecular fragments are represented, each structural fragment must be present an adequate number of times in the training dataset; given the paucity of accurate experimental data, this presents a major problem. Depending upon the size of the fragments used, a GC approach may not be able to distinguish between two different isomers of the same molecule. Bilirubin, which is a metabolite of hemoglobin that can cause jaundice in young babies, provides an interesting example of this problem. The medical cure is the use of phototherapy, which causes a cis-trans isomerism. The trans- product is more soluble in water and is safely excreted via the kidneys from the body.[120] However, the solubility of both molecules is predicted to be the same by many Group Contribution methods.[11] Moreover, application of GC methods for many practical tasks becomes complicated due to the significant number of descriptors which are required for predictions of HFEs for a wide range of solutes.[53,54,87,95,96,115] The problems associated with Group Contribution approaches

have meant that recently the majority of QSPR publications on the prediction of solvation free energy have tended to use other sets of molecular descriptors.[54,95,96,112–115,121] Despite the more advanced nature of these QSPR models as compared to GC approaches, the majority of them still not extrapolate well to molecules dissimilar to those in the training dataset.

2.3.2 'Bottom up' methods: implicit vs explicit solvent

Although cheminformatics/QSPR approaches provide a computationally inexpensive way to predict SFEs and other thermodynamic parameters, these methods do not have molecular-scale model of solvation behind them. As a result, even successful predictions by these methods do not provide chemical/physical insights into the underlying solvation processes; also, as a rule, their applications are limited to selected classes of compounds.

'Bottom up' methods that make predictions starting from a molecule-scale model of the solvation process provide an alternative route for predicting SFEs. Commonly used 'bottom up' methods to calculate SFE may be categorized as either implicit or explicit solvent models.[54,91,105–115] The full theoretical background of these methods has been the subject of a large number of excellent review articles and books (see also above) and will not be recapitulated here. Nevertheless, we provide a brief overview of both methods below in order to place the integral equation theory of molecular liquids in the context of other computational solvent models.

Implicit continuum models consider the solvent to act like a uniform polarizable medium characterized by a dielectric constant, ϵ, into which the solute is embedded in a suitably shaped cavity.[107,122–124] The solvation free energy is decomposed as:

$$\Delta G_{solv} = \Delta G_{elec} + \Delta G_{rep-disp} + \Delta G_{cav}, \tag{14}$$

where the terms on the right hand side are due to polarization of the solute by the solvent environment (ΔG_{elec}), repulsion-dispersion interactions between solute and solvent ($\Delta G_{rep-disp}$), and formation of a cavity in the solvent (ΔG_{cav}). The first term, ΔG_{elec}, can be calculated for

a given molecular geometry by solving either the Generalized-Born or Poisson-Boltzmann equations, which can be done for pure solvents and electrolyte solutions.[109] A wide variety of methods have been proposed to calculate ΔG_{cav} and $\Delta G_{rep-disp}$. A common approach is to combine these terms and represent them by an equation of the form

$$\Delta G_{cav} + \Delta G_{rep-disp} = \gamma SASA + b, \tag{15}$$

where $SASA$ is the solvent accessible surface area of the solute and γ and b are constants obtained by empirical fitting against experimental data for non-polar molecules. The accuracy of this method is questionable since the linear relationship between $\Delta G_{cav} + \Delta G_{rep-disp}$ and $SASA$ only holds for simple organic molecules.

Implicit continuum models ignore the microscopic solvent structure, which precludes an accurate description of specific, short-range solute-solvent interactions that give rise to many of the interesting phenomena of solution chemistry.

Explicit solvent simulations offer a more thermodynamically rigorous approach to calculating solvation free energies.[91,104–106,110,125–131] The standard computational model used for these calculations is a simulation box comprising a solute molecule immersed in a large number of solvent molecules each of which is modeled in atomistic detail. The calculation of solvation free energies requires an exploration of the possible configurations of the molecules in the system (using molecular dynamics or Monte Carlo algorithms) acting under the observed potential energy function (usually evaluated using molecular mechanics force fields).

The calculation of long-range electrostatic interactions in such many-body systems is computationally intensive with the result that large-scale computing resources are required. Solvation free energies can be calculated using explicit solvent simulations by four main groups of methods:[109] (i) thermodynamic integration (TI); (ii) free energy perturbation (FEP); (iii) probability densities; (iv) non-equilibrium work (NEW) methods. Since the free energy (of solvation) is a state function it is (formally) independent of the path used to calculate it. However, convergence of the calculations

(and, consequently, the accuracy of the calculations) depends on the method as well as the energy landscape of the system. A comparative analysis of different methods and their applications can be found here.[54,91,105–115]

A recent series of blind challenges for the calculation of the HFEs of pharmaceutically and environmentally relevant molecules (SAMPL1, SAMPL2, SAMPL3, SAMPL4) has demonstrated the current state of the art of implicit continuum and explicit solvent simulation based approaches.[53,54,95,96] The best predictions were in the range RMSE=2.5-3.5 kcal/mol, which equates to an $\sim$2 log unit error in the related equilibrium physicochemical property (e.g. solubility, pKa, etc). Clearly, these methods are not accurate enough for many practical applications (e.g. modern pharmaceutical research). As the authors of these challenges observe, "Accurate calculations of more complex properties or events will remain over the horizon until these more basic values [HFEs] can be predicted with greater accuracy".[53]

3   Integral Equation Theory (IET) of molecular liquids

The integral equation theory (IET) of molecular liquids has a special place among the implicit solvent methods, since it combines a reasonable level of molecular description of the solute with low computational costs. IET provides information about the solvent density around the solute (in terms of density distribution functions), which is not accessible by continuum solvent methods. Another advantage of IET is its ability to perform single-point SFE calculations with a free energy functional, instead of the series of calculations that are necessary for SFE calculations with most explicit solvent methods. Such fruitful combination of IET's features provoked a range of studies of SFE using the method starting from the final decade of the last century.[107,132–138]

Although qualitative agreement with experimental data is demonstrated in some of these applications, extensive benchmarking has indicated that the standard IET free energy functionals do not give accurate HFEs; indeed experimental and computed values may differ by an order of magnitude. Thus, the original formulation of IET allows only a qualitative correct description of

molecular solvation effects,[18,139] whereas quantitatively accurate prediction of energetic parameters has been a challenge for quite sometimes.[134,138,140] However, recent development of hybrid free energy functionals has allowed IET to approach chemical accuracy in terms of predictions of solvation thermodynamics[2,92,141] All these achievements have brought forth a dramatic growth of interest in this area.

Below we provide a brief overview of the main IET formulae, starting with a description of simple systems (spherical particles interacting via short-range potentials) and then progressing towards IET methods for complex molecular systems. We refer to Refs[142–144] for more detailed descriptions of IET fundamentals.

3.1   The Ornstein - Zernike (OZ) equation for simple fluids

The fundamental relationship in the integral equation theory of liquids is the Ornstein-Zernike (OZ) equation,[147] which defines the *total correlation function* $h(r)$ between a pair of spherical particles 1 and 2 for the case of a homogeneous isotropic fluid with density ρ as:

$$h(r_{12}) = c(r_{12}) + \rho \int c(r_{13})h(r_{32})dr_3, \tag{16}$$

where $c(r)$ is the so-called *direct correlation function* which was first introduced by Ornstein and Zernike. The *pair density distribution function* $g(r)$ can be found from $h(r)$ as $g(r) = h(r) + 1$.
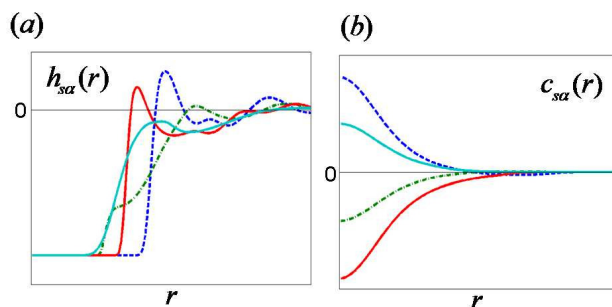
Figure 6: Typical form of the (a) total, and (b) direct site-site correlation functions for different components of a liquid.

$\rho g(r)$ gives the probability to find a particle in the volume element dr located at the distance r from another particle. Therefore, $g(r)$ is also often referred as *pair correlation function* or *pair distribution function*.[142–144,148]

The physical significance of eq 16 is most easily seen if an attempt is made to solve it recursively, i.e. by eliminating $h(r)$ under the integral using the equation:

$$h(r_{12}) = c(r_{12}) + \rho \int c(r_{13})c(r_{32})dr_3 + \\ \rho^2 \int \int c(r_{13})c(r_{34})c(r_{42})dr_3 dr_4 + ... \tag{17}$$

Writing the OZ equation in this form shows that the total pair correlation function between particles 1 and 2 $(h(r_{12}))$ is a sum of the pair direct correlation function between the particles $(c(r_{12}))$ and an indirect correlation function $(\gamma(r_{12}) = h(r_{12}) - c(r_{12}))$ propagated via increasingly large numbers of intermediate particles. Thus, the right-hand side in eq 17 represents an infinite series comprised of "chains" of various direct correlations.[143,148]

One should note that the OZ equation keeps in consideration only the first term in eq (correlation via any particle 3) of the cluster expansion considered here (eq 17), and that in many the following elaborations of the method this limitation to the first term has been silently preserved. We note, however, that high–order correlation functions are important for correct description of inhomogeneous fluids and many important phenomena such as glass transition, wetting, condensation etc.[143–145,148] Therefore, further development of the theory of fluids requires consideration of the higher terms of the cluster expansion.[143–146]

It is worth noting at this point that the OZ equation can be generalized for *mixtures* of spherical particles as:

$$h_{ij}(r_{12}) = c_{ij}(r_{12}) + \sum_{m=1}^{M} \rho_m \int c_{mi}(r_{13})h_{mj}(r_{32})dr_3, \tag{18}$$

where M is the number of components in the mixture. The total and direct correlation functions are illustrated for a typical mixture in Figure 6. At short distances, the total correlation functions reveal sharp peaks and slopes corresponding to solvation shells, whilst at long-range distances the

functions decay weakly with oscillations.

By contrast, the direct correlation functions behave asymptotically with $c(r) \to -\beta u(r)$ as $r \to \infty$; $u(r)$ is the pair interaction potential between the liquid particles, $\beta = 1/k_B T$, k_B is the Boltzmann constant, and T is the absolute temperature. This asymptotic behavior holds true for a wide range of liquids provided that the liquid is not near the critical point.[143,148] Due to the complicated form of the total correlation functions, it is often convenient to operate with indirect correlation functions $\gamma(r_{12})$ which are usually smoother functions.[143]

The OZ equation (eq 16) contains two unknowns functions, $h(r)$ and $c(r)$. In order to solve for these functions, a second equation must be introduced, the so-called "closure relation", which couples $h(r)$ and $c(r)$ with the pair interaction potential $u(r)$. The algebraic form of this closure relation can be written as:[142,143,148]

$$h(r) + 1 = \exp(-\beta u(r) + \gamma(r) + B(r)). \tag{19}$$

The closure relation introduces a so-called "bridge" function ($B(r)$), which is a functional of the indirect correlation function, ($B(r) = B[\gamma(r)]$).

The total and direct correlation functions, $h(r)$ and $c(r)$, respectively, can be found by numerical solution of eq 16 and eq 19 provided that $u(r)$, T, and $B(r)$ are known. Unfortunately, however, an exact expression for $B(r)$ is not currently known. Therefore, a number of different approximations have been introduced for the bridge function. Finding a good bridge parameterization is one of the main problems of the theory of liquids.

Most of commonly used approximate bridge functionals (and hence closure relations) fall into one of two general classes: (i) those based on the local approximation hypothesis; (ii) those that introduce a small number of free parameters into the bridge functional, whose values are obtained self-consistently to give optimal thermodynamic consistency.

The closure relations based on the local approximation hypothesis replace the generally *non-local* bridge functional, $B[\gamma(r)]$, by a local bridge function depending on $\gamma(r)$, $b(\gamma(r))$. It should be noted, however, that so far there is no rigorous theory behind this replacement of the nonlocal

bridge functional by a bridge function. The first and most widely used local approximate clo-sure relations were the Percus-Yevick (PY), the hypernetted chain (HNC) and the mean spherical (MSA) approximations.

The simplest approximation of the bridge functional is the HNC closure[143] that just ignores the bridge, i.e. $B[\gamma(r)] = 0$. Thus, the HNC closure reads:

$$h(r) + 1 = \exp(-\beta u(r) + \gamma(r)). \tag{20}$$

This approximation is mainly used to describe systems with long-range Coulombic interac-tions.

The PY closure[149] is usually applied to describe systems with short-range interactions. It takes the bridge function in the form $B[\gamma(r)] = \ln(1 + \gamma(r)) - \gamma(r)$, so that:

$$h(r) + 1 = \exp(-\beta u(r))(\gamma(r) + 1). \tag{21}$$

The mean spherical approximation (MSA) is characterized by the bridge function $B[\gamma(r)] = \ln(1 + \gamma(r) - \beta u(r)) - \gamma(r) + \beta u(r)$ and has a different form depending on the particle type in the system. For a system consisting of hard core particles, when the asymptotic property of the direct correlation function is assumed to be $c(r) \to -\beta u(r)$ as $r \to \infty$, the MSA has the form:

$$h(r) + 1 = 1 - \beta u(r) + \gamma(r). \tag{22}$$

The MSA can be rewritten for systems comprising of particles with a soft core by splitting the pair interaction potential into an attractive and a repulsive part, $u(r) = u_1(r) + u_2(r)$. The bridge function is taken as the attractive part of the potential, such that the closure can be written as:[143]

$$h(r) + 1 = \exp(-\beta u_1(r))(1 - \beta u_2(r) + \gamma(r)). \tag{23}$$

The HNC and MSA closures were combined in the Kovalenko-Hirata (KH) closure[150] that in

28

the case of OZ equation for simple fluids reads as (see more discussion on this closure for molecular liquids below):

$$h(r) + 1 = \begin{cases} \exp[-\beta u(r) + \gamma(r)] & \text{if } h(r) \leq 0, \\ 1 - \beta u(r) + \gamma(r) & \text{if } h(r) > 0. \end{cases} \tag{24}$$

The Martynov-Sarkisov (MS) closure[151] was obtained by estimation of the density expansion of the bridge functional and from analysis of the bridge functional behavior in the vicinity of the critical point. The bridge functional is taken as $B[\gamma(r)] = \sqrt{1 + 2\gamma(r)} - \gamma(r) - 1$, such that the closure reads:

$$h(r) + 1 = \exp(-\beta u(r) + \sqrt{1 + 2\gamma(r)} - 1). \tag{25}$$

It has been shown that to obtain good local closure approximations it is necessary to explicitly incorporate the pair interaction potential $u(r)$ into the bridge function.[152] Duh, Haymet and Henderson[153,154] used this idea to define a semiempirical (DHH) closure for the OZ equation:

$$B(\Gamma) = -\Gamma^2 \frac{1}{2} \left[1 + \left(\frac{5\Gamma + 11}{7\Gamma + 9} \right) \Gamma \right]^{-1} \tag{26}$$

where $\Gamma(r) = \gamma - \beta u_2(r)$. $u_2(r)$ corresponds to the attractive part of the potential, which is defined by partitioning the potential into a reference (repulsive) part and a perturbation (attractive) part following the method of Weeks et al.[155] The DHH closure quite accurately describes some Lennard-Jones systems. The idea of incorporating a potential term into the bridge-function has also been applied in the construction of the Martynov-Sarkisov-Vompe (MSV) closure.[156]

The second class of approximate closure relations use a small number of free parameters in the bridge functional, whose values are obtained by fitting against experimental data to obtain thermodynamic consistency. Examples of these *self-consistent approximate closures* include the Rogers-Young,[157] the Zerah-Hansen,[158] and the Verlet[159] closures.

The Rogers-Young closure[157] takes the form:

$$B[\gamma(r)] = \ln\left(1 + \frac{\exp\left(f(r)\gamma(r)\right) - 1}{f(r)} - 1\right) - \gamma(r) \tag{27}$$

where $f(r) = 1 - \exp(-ar)$ and a is a self-consistent variable parameter. This closure combines the advantages of the PY closure at small distances ($f(r = 0) = 0$) and of the HNC closure at large distances ($f(r = \infty) = 1$).

The Zerah-Hansen closure[158] combines the HNC approximation at large distances r and the soft core MSA (SMSA) equation at small distances r according to:

$$\begin{aligned} h(r) + 1 = \exp\left(-\beta u_1(r)\right) \times \\ \left(1 + \frac{\exp(f(r)(\gamma(r) - \beta u_2(r))) - 1}{f(r)}\right) \end{aligned} \tag{28}$$

where $u_1(r)$ and $u_2(r)$ are the repulsive and the attractive parts of the pair interaction potential $u(r)$, $f(r) = 1 - \exp(-ar)$, and a is a self-consistent variable parameter.

The Verlet closure[159] has the form

$$B[\gamma(r)] = \frac{1}{2}\frac{\gamma^2(r)}{1 + a\gamma(r)} \tag{29}$$

where a is an empirical parameter. We note that this formula is the basis of the approximation suggested by Duh, Haymet and Henderson, which is described above.

Since it is commonly easier to estimate bridge functionals for simple systems (e.g. hard spheres), the bridge functionals obtained for these simple systems are sometimes used in a renormalized form for more complicated systems. One example is given by the reference (modified) hypernetted chain closure (RHNC or MHNC),[160] which uses $B(r) = B_{HS}(r, \sigma)$ such that the closure relationship reads:

$$h(r) + 1 = \exp\left[-\beta u(r) + \gamma(r) + B_{HS}(r, \sigma)\right] \tag{30}$$

where $B_{HS}(r, \sigma)$ is the bridge function of a hard sphere system, σ is the effective diameter of hard sphere, $u(r) \equiv U(r, \sigma, \epsilon)$ is the potential of the reference system. Malijevsky and Labik have proposed a bridge function $B_{HS}(r, \sigma)$ with coefficients depending on single free parameter, the packing factor $\eta = \pi \rho \sigma^3 / 6$.[161]

3.2   The Molecular OZ (MOZ) equation

The theory outlined in the preceding section is only applicable to atomic liquids. The MOZ equation is the generalization of the Ornstein-Zernike (OZ) equation to non-spherical molecules.[143,148]

For a homogeneous, isotropic fluid, the MOZ equation can be written as:

$$h(\mathbf{r}_{12}, \Theta_1, \Theta_2) = c(\mathbf{r}_{12}, \Theta_1, \Theta_2) +$$
$$\frac{\rho}{\mathbb{Z}} \int c(\mathbf{r}_{13}, \Theta_1, \Theta_3) h(\mathbf{r}_{32}, \Theta_3, \Theta_2) d\mathbf{r}_3 d\Theta_3 \tag{31}$$

where the two-particle total and direct correlation functions are depend on the displacement of the two particles characterized by $\mathbf{r}_{12} = (x, y, z)$ and on the orientation of the two molecules denoted by Θ_1 and Θ_2, which are sets of the three Euler angles $(\psi, \theta, \varphi)(i = 1, 2)$. $\mathbb{Z}$ is equal to 4π or $8\pi^2$ depending on whether the molecule is linear, where the molecule's orientation can be described by two angles (e.g. HCl molecule); or nonlinear, where the molecule's orientation must be described by three angles (e.g. H_2O molecule). Integration is carried out over all positional $(d\mathbf{r}_3)$ and orientational $(d\Theta_3)$ coordinates of molecule 3.

When the concentration of the solute in pure solvent tends to zero (i.e. an infinitely dilute solution), the MOZ equations can be rewritten as three independent equations, operating with *solvent - solvent*, *solute - solvent* and *solute - solute* correlation functions, respectively, which can be solved separately.[143]

In principle, the MOZ equation offers a very powerful method for the calculation of solvent structure. The calculated density correlation functions can be directly compared to experimental observables obtained by structural methods (e.g. neutron scattering). Nevertheless, the MOZ equation is still rarely used to study molecular liquids because of the difficulty of solving the high-

dimensional equations. The simplest approach to this problem involves expanding $h(\mathbf{r}_{12}, \Theta_1, \Theta_2)$, $c(\mathbf{r}_{12}, \Theta_1, \Theta_2)$ and the potential of intermolecular interaction in spherical harmonics.[162] In this form, the expressions become algebraically complex, which leads to a number of problems. Firstly, numerical solutions of the equations converge very slowly. Secondly, it becomes necessary to careful investigate the convergence of the series for each molecular system.

We note, however, that a significant progress in this area was recently made by several groups;[163–166] see more discussion on these methods below.

3.3   Reference Interaction Site Model (RISM): main formulae

To circumvent the problems found in operating with the high-dimensional MOZ equation, several methods have been developed originating from the work of Chandler and Anderson,[167] generally named Reference Interaction Site Models (RISM), that reduce the dimensionality of the original MOZ equations. These methods have been used for a wide range of applications in the chemical sciences[3,7,13,21,92,101,136,141,168–170] and we will mostly focus on them in this review.

Overall, RISM has many attributes that make it an excellent method to complement traditional implicit continuum or explicit solvent simulations (and to replace them in some situations). From one side, the theory allows calculation of solvent density distributions and solvation thermodynamic parameters at significantly lower computational expense than explicit solvent simulations. From another side, the theory may be used to study specific solute-solvent interactions that are not accessible by continuum solvent models. Moreover, the theory is not limited to bulk water solutions and it can be generalized to many different pure and mixed solvent systems.[171–173]

RISM theory has been used to study a wide-variety of physicochemical phenomena: the surface-induced structure of water,[174] salt effects on the solubility of noble gases in water,[175] the conformational stability of peptides in aqueous solution,[176–180] prediction of shape and size distribution of micelles,[181–183] effects of salts[16,184] and alcohol[185] on peptide conformations, electron transfer at water-electrode interfaces,[186] calculation of the partial molar volume of ions[187] and molecules.[188–190] Recently RISM areas of applications have been extended to hydration thermo-

dynamics of proteins,[191–193] supramolecular assemblies,[3,194,195] ion interactions with DNA[196] and proteins,[3,20] computational drug-design,[3,101,197] and solvation effects in ionic liquids[198,199] and in nano-porous materials.[7,200]

3.3.1 1D RISM

In the 1D RISM approach, the 6D MOZ equations are approximated by a set of one-dimensional integral equations, operating with the intermolecular *spherically-symmetric* site-site correlation functions. This means that solute and solvent molecules are treated as sets of sites with spherical symmetry, which allows the OZ equations to be written in terms of correlation functions that depend only on the distance between these sites (see Figure 7).[148] In the RISM theory there are three kinds of site-site correlation functions: intramolecular correlation functions $\omega(r)$, total correlation functions $h(r)$ and direct correlation functions $c(r)$.[143] Due to the spherical symmetry, these functions depend only on the radial distance r between interaction sites.

The benefit of this approach is that the resulting integral equations in the 1D RISM approach can be solved with minimal computational effort for most systems of chemical interest. For example, using current numerical algorithms, the solvation free energy of an average druglike molecule may be calculated in a matter of seconds on a modern personal desktop computer.[2,4,201,202] By comparison, the calculation of solvation free energies for the same molecules by continuum methods takes a few seconds and by explicit solvent simulation takes between hours and days.[54,91,105–115]

In the original 1D RISM method[167] and its extension to dipolar liquids (XRISM[203]) the site-site Ornstein-Zernike (SSOZ) integral equations are solved by combining them with a closure relation. Accurate solution of the SSOZ equation in principle requires integration over an almost infinite series of integrals over high order correlation functions which are practically incomputable. For this reason, an additional approximation is introduced into RISM in the form of a simplified Bridge function in the closure relationship (see below). Despite significant efforts[142,143,153,204,205] the design of appropriate bridge functionals is still currently an open problem in the IET.

The *intramolecular* correlation functions describe the structure of the molecule. We will use
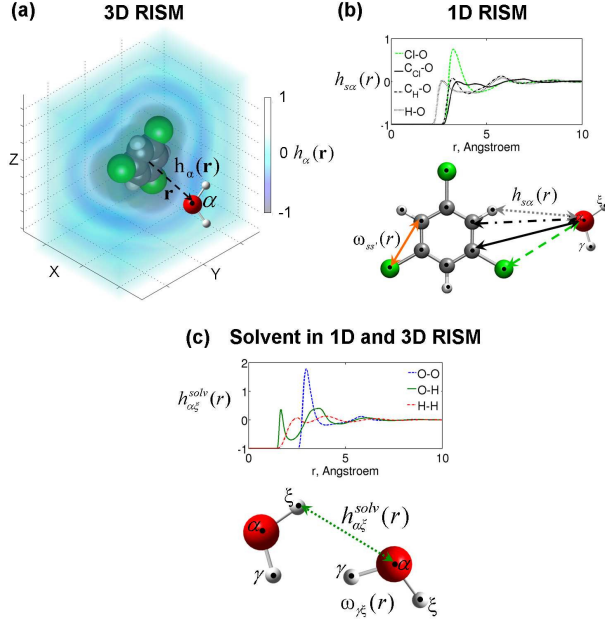
Figure 7: 1D and 3D RISM correlation functions. (a) Intermolecular solute-solvent correlation function $h_\alpha(\mathbf{r})$ computed by 3D RISM for a model solute; (b) Spherically-symmetric correlation functions in 1D RISM: site-site intramolecular ($\omega_{ss'}(r)$) between the site of solute molecule and intermolecular ($h_{s\alpha}(r)$) correlation functions between sites of solute and solvent molecules. The inset plot shows the radial projections of solute site-oxygen water density correlation functions. (c) Solvent-solvent correlations in both 1D and 3D RISM methods: site-site intramolecular correlation functions ($\omega_{\gamma\xi}^{solv}(r)$) and intermolecular correlation functions ($h_{\alpha\xi}^{solv}(r)$) between sites of solvent molecules. The inset shows the radial projections of water solvent site-site density correlation functions: oxygen-oxygen (OO, blue dashed), oxygen-hydrogen (OH, green solid) and hydrogen-hydrogen (HH, red dash-dotted). Reprinted with permission from Ref.[2] Copyright 2011 American Chemical Society.

notation $\omega_{ss'}(r)$ for the intramolecular correlation function between the sites s and s' of the solute molecule and notation $\omega_{\alpha\xi}^{solv}(r)$ for the intramolecular correlation function between sites α and ξ of the solvent molecule. For two given sites, s and s', of one molecule the intramolecular correlation function is written as:

$$\omega_{ss'}(r) = \frac{\delta(r - r_{ss'})}{4\pi r_{ss'}^2}, \tag{32}$$

where $r_{ss'}$ is the distance between the sites and $\delta(r - r_{ss'})$ is the Dirac delta-function.

The *intermolecular* solute-solvent correlations in RISM are described by the pairwise *total* correlation functions $h_{s\alpha}(r)$, and *direct* correlation functions $c_{s\alpha}(r)$ (index s corresponds to the

34

solute sites and index α corresponds to the solvent site). The total correlation functions $h_{s\alpha}(r)$ can be expressed via the radial distribution function (RDF) as:

$$h_{s\alpha}(r) = g_{s\alpha}(r) - 1, \tag{33}$$

where $g_{s\alpha}(r)$ is RDF of sites α of the solvent molecule around the site s of the solute. The model also uses the solvent-solvent total correlation functions $h_{\alpha\xi}^{solv}(r)$ which give the distribution of sites ξ of solvent molecules around the site α of a selected single solvent molecule.

The direct correlation functions $c_{s\alpha}(r)$ are related to the total correlation functions via the set of 1D RISM integral equations:[143]

$$
\begin{aligned}
h_{s\alpha}(r) = \sum_{s'=1}^{M} \sum_{\xi=1}^{N} \int_{R^3} \int_{R^3} \omega_{ss'}(|\mathbf{r_1} - \mathbf{r'}|) \times \\
c_{s'\xi}(|\mathbf{r'} - \mathbf{r''}|) \chi_{\alpha\xi}(|\mathbf{r''} - \mathbf{r_2}|) d\mathbf{r'} d\mathbf{r''}
\end{aligned}
\tag{34}
$$

where $r = |\mathbf{r_1} - \mathbf{r_2}|$ and $\chi_{\alpha\xi}(r)$ are the bulk solvent susceptibility functions, M and N are number of sites in solute and solvent, correspondingly.

The mutual correlations between sites located on molecules in bulk solvent are described by the solvent susceptibility function $\chi_{\xi\alpha}(r)$. This function can be obtained from the solvent site-site total correlation functions ($h_{\xi\alpha}^{solv}(r)$) and the 3D structure of a single solvent molecule (intramolecular correlation function, $\omega_{\xi\alpha}^{solv}(r)$. See Figure 7, c)):[143,168]

$$\chi_{\xi\alpha}(r) = \omega_{\xi\alpha}^{solv}(r) + \rho h_{\xi\alpha}^{solv}(r) \tag{35}$$

where ρ is the bulk number density of the solvent (here and after we imply that each molecule site is unique in the molecule, so that $\rho_{\alpha} = \rho$ for all α if not specified otherwise).

To make eq 34 complete, $N \times M$ *closure* relations are introduced:

$$h_{s\alpha}(r) = \exp(-\beta u_{s\alpha}(r) + \gamma_{s\alpha}(r) + B_{s\alpha}(r)) - 1$$
$$s = 1, \ldots, M \tag{36}$$
$$\alpha = 1, \ldots, N$$

where $u_{s\alpha}(r)$ is a pair interaction potential between the sites s and α, $B_{s\alpha}(r)$ are site-site bridge functions (see below).

The interaction potential is commonly represented by a long range electrostatic term $u_{s\alpha}^{el}(r)$ and a short-range *Lennard-Jones* (LJ) term $u_{s\alpha}^{LJ}(r)$ as:

$$u_{s\alpha}(r) = u_{s\alpha}^{el}(r) + u_{s\alpha}^{LJ}(r),$$
$$u_{s\alpha}^{el}(r) = \frac{q_s q_\alpha}{r}; \tag{37}$$
$$u_{s\alpha}^{LJ}(r) = 4\epsilon_{s\alpha}^{LJ} \left[\left(\frac{\sigma_{s\alpha}^{LJ}}{r} \right)^{12} - \left(\frac{\sigma_{s\alpha}^{LJ}}{r} \right)^6 \right],$$

where q_s and q_α are the partial electrostatic charges of the corresponding solute and solvent sites, $\epsilon_{s\alpha}^{LJ}$ and $\sigma_{s\alpha}^{LJ}$ are the LJ solute-solvent interaction parameters.

In general, the exact bridge functions $B_{s\alpha}(r)$ in eq 36 are practically uncomputable and one needs to use some approximation.[142,143,153] As in the case of simple fluids considered above, the most straightforward and widely used model is the HNC approximation, which sets $B_{s\alpha}(r)$ to zero.[204] The corresponding closure then reads:

$$h_{s\alpha}(r) = \exp\left[-\beta u_{s\alpha}(r) + \gamma_{s\alpha}(r)\right] - 1.$$
$$s = 1, \ldots, M \tag{38}$$
$$\alpha = 1, \ldots, N$$

However, due to the uncontrolled growth of the argument of the exponent the use of the HNC closure can lead to a slow convergence rate, and in many cases even divergence of the numerical solution of RISM equations.[206,207] One way to overcome this problem is to linearize the exponent

in eq 38 when its argument is larger than a certain threshold constant C:

$$h_{s\alpha}(r) = \{ \begin{cases} \exp(\Xi_{s\alpha}(r)) - 1 & \Xi_{s\alpha}(r) \leq C, \\ \Xi_{s\alpha}(r) + \exp(C) - C - 1 & \Xi_{s\alpha}(r) > C, \end{cases} \tag{39}$$

where $\Xi_{s\alpha}(r) = -\beta u_{s\alpha}(r) + \gamma_{s\alpha}(r)$. The linearized HNC closure for the case $C = 0$ was proposed by Hirata and Kovalenko[208] and has since been referred to as either the partial-linearized hypernetted chain (PLHNC) closure or the KH closure. When C goes to infinity, the PLHNC closure (eq 39) becomes the HNC closure (eq 38).

As a systematic approach to treat the HNC convergence problems, Kast and Kloss proposed to use a partial series expansion of order n (PSE-n) of the HNC closure:[207]

$$h_{s\alpha}(r) = \begin{cases} \exp(\Xi_{s\alpha}(r)) - 1 & \text{if } \Xi_{s\alpha}(r) \leq 0, \\ \sum_{i=0}^{n} (\Xi_{s\alpha}(r))^i / i! - 1 & \text{if } \Xi_{s\alpha}(r) > 0. \end{cases} \tag{40}$$

The PSE-n closures interpolate between the KH and HNC closures: setting $n = 1$ results in the KH closure; $n \to \infty$ gives the HNC closure and its convergence issues. PSE closures of order 3 or 4 achieve a good balance between these limit cases: they have been found to provide good numerical convergence and they give results that well approximate calculations with the HNC closure.[207] Due to these properties PSE-3 and PSE-4 as well as their 3D versions (see below) have been extensively applied recently for a variety of polar and charged systems.[138,170,173,196,209]

One drawback of the RISM and XRISM theories[203,210–213] is that they give a trivial result for the static dielectric constant of a polar liquid.

$$\epsilon_{RISM} = 1 + 4\pi\beta\rho\mu^2/3 \tag{41}$$

where μ is the dipole moment of a solvent molecule and ρ is the solvent number density. One solution to this problem is offered by ARISM, which applies a coefficient, A, to the interaction potential

to scale electrostatic interactions and to ensure that the desired dielectric constant is obtained.[214] Although this approach can be used to study salts at infinite dilution, it leads to several inconsistencies at finite salt concentrations.[215] Currently, the most widely-adopted method to model such finite concentration salt solutions is the dielectrically consistent RISM (DRISM), which allows the desired dielectric constant to be computed under a wide-range of thermodynamic conditions. In the DRISM formalism, the RISM closure is modified by an additional term similar to a bridge function (given here for the HNC closure):

$$
\begin{aligned}
h_\alpha(r) &= \exp(-\beta u_\alpha(r) + \gamma_\alpha(r) + b_\alpha(r)) - 1 \\
\alpha &= 1, \ldots, N
\end{aligned}
\tag{42}
$$

where $b_\alpha(r)$ is a bridge correction included to account for dielectric screening effects that are poorly modelled in the original site-site RISM theory. A wide-variety of different closures are compatible with DRISM.[209,216] The ion-ion correlation functions obtained using DRISM are both consistent with the behaviour of solvent molecules and dielectrically correct. This effect arises by indirect coupling of the correlations with solvent molecules since the ion-solvent and ion-ion interaction potentials are not altered a priori. The XRISM, ARISM, and DRISM methods have been used to study aqueous ionic solutions in both the infinite dilution regime[137,171,187,213,217–221] and over a range of finite salt concentrations.[175,215,222–229] These studies have used a number of different water and ion models at both ambient conditions[141,221] and at extreme temperatures,[225,226,228,229] pressures,[228,229] and concentrations.[227]

3.3.2   3D RISM

1D RISM does not properly take into account the spatial correlations of the solvent density around the solute due to the approximations inherent to the method. Therefore, in 1996 Beglov and Roux proposed a three-dimensional extension of RISM – 3D RISM.[230,231] In the 3D RISM method, the 6D solute-solvent MOZ equation is approximated by a set of 3D integral equations via partial integration over the orientational coordinates.[230,231] In the case of the 3D RISM method, instead

of one 6D MOZ equation one has to solve N solvent 3D equations, which is computationally more feasible. 3D RISM is commonly considered to be more useful than 1D RISM for modeling complex macromolecules, such as proteins, where knowledge of the solvent structure at local surface sites is important in interpreting biological function. The computational expense required to solve the 3D RISM equations is typically of the order of 100-fold greater than that required to solve the 1D RISM equations for a given molecular system.

The 3D RISM equations operate with the intermolecular *solvent site - solute* total correlation functions $h_\alpha(\mathbf{r})$, and direct correlation functions $c_\alpha(\mathbf{r})$ (Figure 7):

$$h_\alpha(\mathbf{r}) = \sum_{\xi=1}^{N} \int_{R^3} c_\xi(\mathbf{r} - \mathbf{r}') \chi_{\xi\alpha}(|\mathbf{r}'|) d\mathbf{r}',$$
$$\alpha = 1, \ldots, N$$

(43)

where $\chi_{\xi\alpha}(r)$ is the bulk solvent susceptibility function, ξ, α denotes the index of sites in a solvent molecule, , and N is the number of sites in a solvent molecule.

The solvent susceptibility functions in eq 43 can be calculated once for a given solvent at certain thermodynamic conditions by 1D RISM (see eq 35) and then they enter the 3D RISM equations as known input parameters.

To make eq 43 complete, N *closure* relations are introduced:

$$h_\alpha(\mathbf{r}) = \exp(-\beta u_\alpha(\mathbf{r}) + \gamma_\alpha(\mathbf{r}) + B_\alpha(\mathbf{r})) - 1$$
$$\alpha = 1, \ldots, N$$

(44)

where $u_\alpha(\mathbf{r})$ is the 3D interaction potential between the solute molecule and α site of solvent, $B_\alpha(\mathbf{r})$ are bridge functions.

The 3D interaction potential between the solute molecule and α site of solvent ($u_\alpha(\mathbf{r})$, see eq 44) is estimated as a superposition of the site-site interaction potentials between solute sites and the particular solvent site ($u_{s\alpha}(r)$), which depend only on the absolute distance between the two

sites:

$$u_\alpha(\mathbf{r}) = \sum_{s=1}^{M} u_{s\alpha}(|\mathbf{r}_s - \mathbf{r}|) \tag{45}$$

where $\mathbf{r}_s$ is the radius-vector of solute site (atom), M is the number of sites in the solute molecule.

When the bridge functions $B_\alpha(\mathbf{r})$ are all set to zero it gives the 3D HNC closure[232] that reads as:

$$h_\alpha(\mathbf{r}) = \exp(-\beta u_\alpha(\mathbf{r}) + \gamma_\alpha(\mathbf{r})) - 1$$
$$\alpha = 1, \ldots, N. \tag{46}$$

Consequently, a 3D version of the PSE-n expansion[207] (eq 40) of the 3D HNC closure takes the form:[196,207]

$$h_\alpha(\mathbf{r}) = \begin{cases} \exp\left(\Xi_\alpha(\mathbf{r})\right) - 1 & \text{if } \Xi(\mathbf{r}) \le 0, \\ \sum_{i=0}^{n} (\Xi_\alpha(\mathbf{r}))^i / i! - 1 & \text{if } \Xi_\alpha(\mathbf{r}) > 0, \end{cases} \tag{47}$$

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

The PLHNC closure in the case of 3D RISM reads as:

$$h_\alpha(\mathbf{r}) = \begin{cases} \exp(\Xi_\alpha(\mathbf{r})) - 1 & \Xi_\alpha(\mathbf{r}) \le C, \\ \Xi_\alpha(\mathbf{r}) + \exp(C) - C - 1 & \Xi_\alpha(\mathbf{r}) > C. \end{cases} \tag{48}$$

We note that in the literature the combination of the PLHNC (or KH when $C = 0$) closure relations (eq 48) and the 3D RISM equations (eq 43) is sometimes referred to as 3D RISM/PLHNC (or 3D RISM/KH) theory,[101,168] but for succinctness we will use 3D RISM instead.

4   RISM: practical aspects

4.1   Numerical algorithms for solving RISM equations

Although RISM is a computationally inexpensive technique compared to explicit solvent simulation, reducing the time required to solve the RISM equations is still of significant importance, es-

(a) 3D RISM

3D box

buffer

$h_\alpha(\mathbf{r})$

1

0

-1

spacing
(grid step)

grid points

(b) 1D RISM

$h_{s\alpha}(r)$

s α

$h_{s\alpha}(r)$

0

r

grid points   grid step

cutoff distance
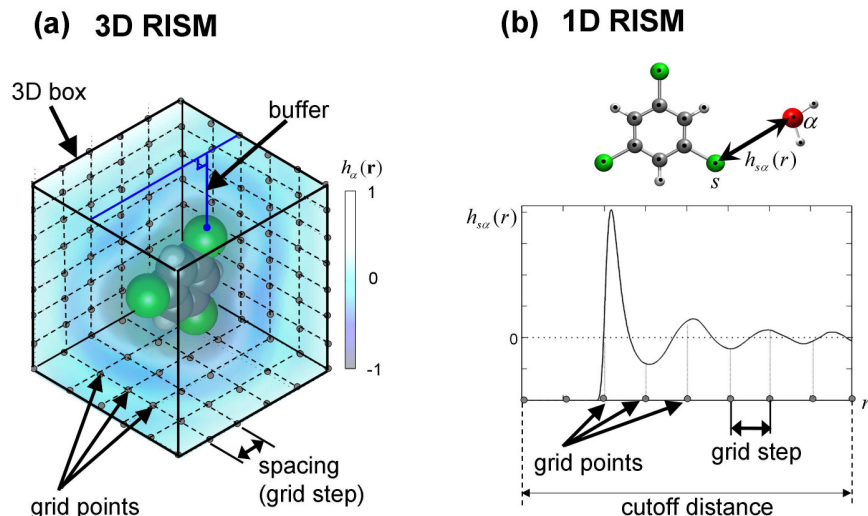
Figure 8: (a) Representation of the 3D-grid box in calculations of total correlation function ($h_\alpha(r)$, where α is the solvent site) within the 3D RISM. Grid points are shown only at the edges of the 3D-box. (b) Representation of a grid in calculations of total site-site correlation function ($h_{s\alpha}(r)$, where s and α are solute and solvent sites, accordingly) within the 1D RISM. Number of grid points and values of grid step and cutoff distance are specified in the text. Republished with permission from Ref.[233] Copyright 2011.

pecially for applications in which large numbers of calculations are required, e.g. high-throughput virtual screening, or when RISM is used in combination with Monte Carlo or molecular dynamics simulations.[1,234–237] Even for simple examples such as an isotropic liquid the RISM requires a non-trivial *numerical* solution of a system of OZ-type integral equations.[148] The complexity of the solution increases dramatically with the number of different interacting sites of the system.[1,143,238]

The most straightforward algorithm to solve the OZ-type equations is the *Picard algorithm* which is based on a successive substitution scheme (this method is sometimes called "direct iteration method"). This technique is easy to implement but it suffers from poor convergence.[143,221,239]

The more advanced methods that are most commonly used to solve OZ-type equations can be roughly classified into four general domains. In the first category there are methods based on Picard iterations with techniques like vector extrapolation or iterative subspace extrapolation used to improve the convergence.[1,240–242] These methods are closely related to the method of modified direct inversion of the iterative subspace (MDIIS) that is actively used by the RISM community.[243]

Maruyama and Hirata further developed this approach by using modified Anderson method to accelerate 3D RISM calculations on Graphics Processing Units (GPUs).[5] This work shows that the modified Anderson method on GPUs provides greatly accelerated convergence of 3D RISM calculations when compared to the MDIIS method on Central Processing Units (CPUs).[5]

In the second category there are hybrid Newton-Raphson/Picard iteration methods. We note that the Newton-Raphson (NR) method itself is not typically an efficient way to solve IET equations because it requires the calculation and inversion of a large Jacobian matrix. In this hybrid methods, the dimensionality of the Newton-Raphson part is reduced by expansions of the IET equations in basis sets of roof functions,[244] plane waves[245–247] or wavelets.[221,248,249]

In the third category are procedures for solving the OZ equation using matrix-free iterative Krylov or Newton-GMRES solvers.[250,251]

In the final category there are methods[4,201,239] that are based on the so-called mutligrid approach.[252–256] These methods use a combination of coarse and fine grid iterations to accelerate convergence. A thorough benchmark of the performance of the multigrid method for SFE calculations and comparative analysis of different numerical methods for 1D and 3D RISM equations can be found here.[4] See also Appendix 3 for a typical model set-up for HFE calculations of organic molecules by RISM.

4.2   Hybridization of RISM with other methods (QM and MM)

The Nobel Price in Chemistry 2013 to Karplus, Levitt and Warshel for the development of QM/MM methods for studying molecular systems reflected the popularity of multi-scale concepts, where different parts of the system are treated at different levels of complexity.[257–266] The current status of QM/MM methods is reviewed elsewhere.[267–269] In the framework of this concept RISM (and IET in general) can be linked to other methods. The advantage of RISM is that, as we show below, it can be linked with methods from both quantum mechanics and molecular mechanics domains. In this section we discuss several examples of incorporating RISM solvation models in a multiscale description of complex molecular systems.

4.2.1 RISM–QM

Up to date there have been already developed several hybrid modeling techniques that couple RISM with different Quantum Mechanics (QM) methods (RISM–QM) to study effects of solvent on the solute electronic structure.

The RISM–SCF method proposed by Ten-no et al[270] couples the extended RISM theory with the Hartree-Fock equation to allow the solvation effect on the electronic structure of a solute to be calculated. The variational equation subject to the orthonormalization constraint in the RISM–SCF approach was originally given by Ten-no et al[270] as:

$$\sum_i \langle \delta\phi_i \mid F_i - f_i \sum_\lambda V_\lambda b_\lambda \mid \phi_i \rangle = 0 \tag{49}$$

where f_i is the occupation number of orbital ϕ_i, b is a population operator for solute site λ, V_λ is the electrostatic potential induced by solvent on site λ, and F is the Fock operator for the intramolecular portion of solute. The term $F_i^{solv} = F_i - f_i \sum_\lambda V_\lambda b_\lambda$ is commonly denoted the "solvated" Fock operator. The standard self-consistent molecular orbital calculation is coupled with RISM through site-site Coulombic interactions. The electrostatic potential induced by solvent on site λ of solute can be calculated from the RISM total correlation function $h_{\lambda\alpha}$ as:

$$V_\lambda = \rho \sum_\alpha q_\alpha \int_0^\infty 4\pi r^2 \frac{h_{\lambda\alpha}(r)}{r} dr \tag{50}$$

where q_α is the atomic partial charge on solvent site α. The RISM equations themselves are solved using the atomic partial charges calculated from the wave-function, which are normally obtained by least-squares fitting to reproduce the electrostatic potential at grid points outside the solute. The partial charges on solute site λ are expressed from the population operator, b_λ, by summing over all the contributions due to electrons and nuclei:

$$q_\lambda = q_\lambda^{(N)} - \sum_i f_i \langle \phi_i \mid b_\lambda \mid \phi_i \rangle \tag{51}$$

The electronic and solvent structure of the system may then be solved self-consistently using the cycle: (a) calculate partial charges on solute atoms from current wave-function; (b) compute the solvent distribution from RISM using the current partial charges; (c) calculate the electrostatic potential on each solute atom using eq 50; (d) recompute the wave-function for the next step of the SCF calculation using the new solvated Fock operator. The original RISM–SCF method was extended by Sato et al[271] to derive the variational conditions of the multiconfigurational self-consistent field (RISM–MCSCF) method and the expression for the analytical gradient of the free energy with respect to the solute nuclear coordinates.

Although the original RISM–SCF approach has been used widely, the least-squares fitting procedure used to calculate atomic partial charges leads to the following drawbacks: (i) the calculated partial charges depend on the choice of grid points; (ii) the evaluation of point charges is an ill-posed problem for some solutes or solute conformations, e.g. molecules with buried atoms; (iii) the point charge representation is a poor model of the electron distribution. To address these problems, Yokogawa et al[272] proposed the RISM–SCF–SEDD procedure, where SEDD stands for spatial electron density distribution. In this approach, the electron density of the solute is expanded in a set of spherical auxiliary basis functions, following the procedure of Gill et al.[273]

The RISM–SCF and RISM–SCF–SEDD procedures use the site-site Ornstein-Zernike equation (1D RISM) to model the influence of solvent on the electronic structure of the solute. A natural extension of these methods is to couple the more theoretically advanced 3D RISM approach with quantum mechanics calculations, which has been done for both *ab initio* molecular orbital calculations[274] and Kohn-Sham density functional theory.[10,205] The principle benefit of these methods is that unlike the original RISM–SCF they allow the 3D spatial distribution functions of solvent molecules around solute to be calculated. The orientational distribution of solvent molecules around solute can be obtained within the RISM–QM framework using the MOZ-SCF method proposed by Yoshida and Kato.[162]

In the combined Density Functional Theory (DFT) and 3D RISM method (DFT/3DRISM) proposed by Gusarov et al[10] and implemented in the Amsterdam Density Functional (ADF) program,

the electronic structure of the solute is calculated from the self-consistent KS-DFT equations modi-

fied to include the presence of solvent. The total system of the solute and solvent has the Helmholtz

free energy defined as

$$A\left[n_e(\mathbf{r}), \{\rho_\gamma(\mathbf{r})\}\right] = \\ E_{solute}\left[n_e(\mathbf{r})\right] + \Delta\mu_{solv}\left[n_e(\mathbf{r}), \{\rho_\alpha(\mathbf{r})\}\right] \tag{52}$$

where E_{solute} is the electronic energy of the solute consisting of the standard components, $\Delta\mu_{solv}$

is the excess chemical potential of solvation coming from the solute-solvent interaction and sol-

vent reorganization due to the presence of the solute, $n_e(\mathbf{r})$ is the electron density distribution,

and $\{\rho_\alpha(\mathbf{r})\}$ are the classical density distributions of interaction sites $\alpha = 1, ..., s$ of the solvent

molecule(s). The solute energy is determined by the standard KS- DFT expression.[10]

The self-consistent Kohn-Sham equation for the case of a solute in solution which can be

obtained by minimization of the free energy functional subject to the standard normalization con-

ditions, is given by:

$$\left[-\tfrac{1}{2}\Delta + v_i(\mathbf{r}) + v_h(\mathbf{r}) + v_{xc}(\mathbf{r}) + v_{solv}(\mathbf{r})\right]\psi_j(\mathbf{r}) \\ = \epsilon_j\psi_j(\mathbf{r}) \tag{53}$$

where $v_i(\mathbf{r})$ comprises the external potential and the nuclear attractive potential, $v_h(\mathbf{r})$ is the

Hartree potential, $v_{xc}(\mathbf{r})$ is the exchange-correlation potential, ϵ_j is the energy of the Kohn-Sham

orbital, $\psi_j(\mathbf{r})$. The addition term, v_{solv}, is the solvent potential, which is computed using 3D RISM

as:

$$v_{solv}(\mathbf{r}) = \frac{\delta\Delta\mu_{solv}\left[n_e(\mathbf{r}), \{\rho_\gamma(\mathbf{r})\}\right]}{\delta n_e(\mathbf{r})}. \tag{54}$$

In 2008 Kloss, Heil and Kast developed the embedded cluster reference interaction site model

(EC-RISM)[275] that uses a self-consistent procedure for the determination of the electronic structure

of the solute and the embedding solvent. The main idea of the method is that solvent distribution

functions are computed via 3D RISM and then the resulting solvent charge distribution is mapped

onto a set of discrete background point charges which interact with the solute electronic structure which in turn influences the solvent response. It was shown that for a small organic molecule the procedure converges in less than 10 iteration cycles.[275] The EC-RISM approach has several advantages for practical applications of RISM-QM because (in the words of the authors of Ref.[275]): "The EC-RISM framework is easily coupled to an existing quantum chemistry code that allows the specification of discrete background charges. The iterative cycles can be controlled entirely by scripting languages, do not require modifications to the quantumchemical program, and converge quite quickly. The methodology works seamlessly with variational and non-variational theories, both for the integral equation and for the quantum chemistry part. As in the RISM-SCF approach, the solvent granularity and H-bonding are adequately accounted for. The framework can directly be applied to heterogeneous systems such as a reacting, solvent-exposed species in the active site of an enzyme."[275]

The use of RISM based methods to model solvent in electronic structure calculations is still a relatively new field. Nevertheless there have been reported significant successes of applying RISM–QM modeling methods to various problems in different areas of chemical sciences.

Sato et al[271] used the RISM–SCF and RISM–MCSCF methods to study the influence of aqueous solvent on the relative stabilities of *cis-* and *trans*-isomers of 1,2-difluoroethylene. While the *trans*-isomer was correctly identified to be more stable in both gas and solution, a relatively larger stabilising effect due to solvent was observed for the *cis*-isomer. Electron correlation was shown to enhance the solvation for the studied systems. Sato et al[276] compared the wave-functions of all species in the reaction $Cl^- + CH_3Cl \rightarrow ClCH_3 + Cl^-$ calculated using RISM and PCM solvent models at the RHF/6-311G* level of theory. The potentials of mean force were very similar for the two methods, but the electronic distortion due to solvation obtained by PCM was slightly smaller than that obtained by RISM–SCF. Using the energy partitioning scheme proposed by Nakai et al[277] and Sato et al,[276] the authors demonstrate that the $Cl-CH_3$ bond is more polarized in the wave-function obtained using PCM, while the C-H bond is important in the wave-function obtained by RISM–SCF. Vchirawongkwini et al[278] used RISM–SCF–SEDD calculations at the

HF/6-311+G* and B3LYP/6-311+G* level of theory to illustrate that the experimentally observed breaking of symmetry of carbonate anions (from D_{3h} to C_{3v}) in aqueous solution is caused by bulk solvent rather than specific solvent molecules. The RISM–SCF method was applied by Ten-no et al[279] to study the solvent effect on carbonyl compounds (formaldehyde, acetaldehyde, acetone and acrolein). The vertical transition energies ($n \rightarrow \pi, \pi \rightarrow \pi*, \sigma \rightarrow \pi*$) of formaldehyde were found to be shifted relative to those *in vacuo* due to presence of solvent, with the $\pi \rightarrow \pi*$ shift in acrolein predicted to be different from those in the other complexes. As would be expected, the dipole moment of the solute was enhanced by aqueous solvent.

Sato et al[280] proposed a method to model the reorganization of solvent that occurs when the charge of an organic compound is changed. The method uses RISM-SCF coupled with a procedure to evaluate nonequilibrium solvation free energy developed by Chong et al.[281] The process of charging N,N-dimethylaniline (DMA $\rightarrow$ DMA$^+$) in acetonitrile solvent, a popular system in studies of the charge-transfer reaction, was chosen as a model system. It was shown that the calculated energy of the charge-transfer process was in reasonable agreement with the experimental data. Moreover, it was shown that the reorganization of solvent could be partitioned into contributions from the atoms of the solute molecule. Analysis of the model system revealed that the energy is not much effected by solute geometry while the individual atomic contribution is significantly changed, especially by the motion of the amino-group.

Combinations of the 3D RISM and quantum DFT methods have been used to model a wide-variety of different chemical systems and phenomena, including solvation effects on conformational equilibria,[168] molecular structure in ionic liquids,[282] the electronic structure of metals,[283] the aggregation of asphaltenes,[284–286] the behaviour of polyester ionomers,[287] and the absorption spectroscopy of nickel(II) and vanadyl porphyrins relevant to bitumen and crude oils.[288]

The EC-RISM approach was successfully applied for prediction of tautomer ratios,[21] studying solvation effects on NMR chemical shifts[170] and prediction of acidity constants of organic molecules in dimethyl sulfoxide (DMSO) solution.[173]

4.2.2 RISM–MM

Liquid-phase environment considerably effects the equilibrium distribution of solute conformations. Accounting for the solvent effect is particularly important in investigations of flexible biomolecules (e.g. peptides, proteins, etc).[3,7,24,25,143,289] To study these effects, 1D and 3D RISM have been coupled with conformational sampling methods (e.g. Monte Carlo (MC) simulated annealing, a multicanonical algorithm in Molecular Dynamics (MD), etc.).[3,7,24–27,143,289] One advantage of this approach is that 1D and 3D RISM can simulate very dilute (e.g. electrolyte or drug-like molecule) solutions,[13,196,209,290] which are difficult to study satisfactorily with explicit solvent simulation methods. The combination of a conformational sampling method with RISM allows a detailed description of the structure and flexibility of molecules in solution.[3,7,24–27,143,289]

The hybridization methodology used in combined RISM and Monte Carlo approaches (RISM–MC) is based on the following considerations. In the case of infinitely diluted solution, the potential energy of the system can be represented as a sum of three contributions:

$$U(\mathbf{r}) = U_A + U_B + U_{AB}, \tag{55}$$

where $U_A \equiv U_A(\mathbf{r}_A)$ and $U_B \equiv U_B(\mathbf{r}_B)$ are the potential energies of the isolated solute molecule (subsystem A) and solvent molecules (subsystem B), respectively, where their interactions are characterized by $U_{AB} \equiv U_{AB}(\mathbf{r}_A, \mathbf{r}_B)$; $\mathbf{r}_A$ and $\mathbf{r}_B$ are configurations of the subsystems. The potential of mean force of the system is given by:

$$\Psi(\mathbf{r}) = U_A(\mathbf{r}) + \Delta\Psi(\mathbf{r}), \tag{56}$$

where $\Delta\Psi(\mathbf{r})$ is the solvent-inducted potential. It determines the influence of solvent molecules on the conformation of the solute molecule in solution. For a $N-$site solute molecule this potential can be written as:[289]

$$\Delta\Psi(\mathbf{r_1}, \ldots, \mathbf{r_N}) = \sum_{s<s'}^{N} \Delta\Psi_{ss'}(r), \tag{57}$$

48

where s and s' are two solute's sites, r is the distance between the sites, $\Delta\Psi_{ss'}(r)$ is the corresponding two-side (pair) component. There are several works on determination of the form of the pair solvent-mediated potential $\Psi_{ss'}(r)$.[204,291,292] Chandler et al[291] obtained it as:

$$\Delta\Psi_{ss'}(|\mathbf{r}_s - \mathbf{r}_{s'}|) = -k_B T \int \int c_{s\alpha}(|\mathbf{r}_s - \mathbf{r}_\alpha|) \times$$
$$\chi_{\alpha\beta}(|\mathbf{r}_\alpha - \mathbf{r}_\beta|) c_{\beta s'}(|\mathbf{r}_\beta - \mathbf{r}_{s'}|) d\mathbf{r}_s d\mathbf{r}_{s'} \tag{58}$$

The main bottleneck in this approach is that the potential of mean force $\Psi(\mathbf{r})$, which determines the potential field acting on the solute molecule in solution, depends not only on the nature of the solute-solvent interactions and liquid-phase properties, but also on the conformation of the solute molecule. This condition requires a self-consistent determination of $\Psi(\mathbf{r})$ such that the equilibrium conformation of the solute molecule $\mathbf{r}_A$ and the effective potential $\Psi(\mathbf{r})$ are mutually-dependent.

One of the earliest publications about the hybrid RISM – Monte Carlo (RISM–MC) approach was published by Khalatur in 1995.[293] The method was tested on a flexible polymer chain in condensed phase imitated by Lennard-Jones particles with purely repulsive interactions between the particles and chain beads. Unfortunately, the article[293] is not easy to access for many in the international scientific society. The self-consistent scheme of RISM–MC calculations is available in a later work of Khalatur and Khokhlov.[289] The scheme contains one preparation step and three main steps (see Appendix 1). A distinctive feature of this scheme is that it operates with an intramolecular correlation function that is averaged over several different conformations (Step 1, c). It leads to a decrease in the number of RISM-steps and speeds-up the convergence of the algorithm. Khalatur and Khokhlov tested the combination of 1D RISM with MC simulated annealing on the flexible polymer chain in a solvent imitated by Lennard-Jones particles.[289]

In 1998, Khalatur et al[294] applied this self-consistent scheme to investigate a collapse of flexible polymer chain surrounded by colloidal particles (modeled by LJ-particles). They tested 1D RISM with two closure relations (HNC and Percus-Yevick, PY[149]). It was found that the results obtained by the HNC closure are in better agreement with the data obtained previously by lattice simulations.

In the same year, Kinoshita, Okamoto, and Hirata[177] proposed a similar self-consistent scheme

for 1D RISM – MC calculations and applied it to investigate conformations of Met-enkephalin in two different solvents: (i) the SPC/E water and (ii) a simple, repulsive potential system. They initiated the calculations from five different conformations of the solute and found that the most stable gas-phase conformation is also the most probable in water. The authors revealed that the simulation located the minimum energy conformation more quickly in solution than in gas-phase. They concluded that water decreases the number of probable solute conformations in solution.

In 2002, Kast co-workers[236] extended the hybrid 1D RISM – MC technique to study complexation thermodynamics (see Appendix 2). The proposed flowchart was applied for investigation of 18-crown-6 complexes with alkali metal ions in methanol and acetonitrile. The main computational advantage of the proposed scheme is the use of the incremental chemical potential approximation that relates a change of the solvent-induced chemical potential to a change of conformation starting from one initial set of 1D RISM equations. It is, therefore, not necessary to recompute the explicit RISM solution for every new configuration, thus accelerating the computations substantially.

As was noted by Kinoshita et al,,[177] the RISM–MC technique is only appropriate for investigations of relatively small compounds and short peptides, whereas long peptide chains and proteins (which have considerably larger conformational flexibility) should be studied with the multicanonical approach, which allows very efficient sampling of phase space. In brief, the multicanonical algorithm is constructed in such a way that a solute configuration with any energy has equal probability (in other words, all energies have equal weight) and a 1D random walk in energy space is realized. As a result, the system can overcome any energy barrier.

The first attempt to combine the multicanonical approach and 1D RISM was performed in 2000 by Hirata and co-workers.[295] Met-enkephalin in aqueous solution was chosen as an example system. The authors showed that the proposed technique covers much wider configurational space than the canonical simulations. Moreover, the method allows one, not only to locate the energy global minimum, but also to calculate the ensemble average of any physical quantity for a wide range of temperatures. The authors calculated averages for the total energy, its component terms, end-to-end distance, and the distributions of the backbone dihedral angles.

Later on Miyata and Hirata combined the MD multicanonical method with the 3D RISM theory.[23] As a benchmark model they chose an aqueous solution of acetylacetone. No intramolecular hydrogen bonds were found within the solvated solute molecule, whereas these bonds were observed in gas-phase calculations. In turn, the most stable solute conformation found by the 3D RISM–MD method had no intramolecular hydrogen bonds. The explanation provided by the authors was that the energy required to break hydrogen bonds is compensated by the formation of stable solute – solvent hydrogen bonds.

Recently Omelyan and Kovalenko developed an efficient computational scheme for handling of solvation forces in the multiscale method of multiple time step molecular dynamics (MTS-MD) of a biomolecule steered by effective solvation forces obtained from the 3D-RISM calculations of SFE.[27] The approach is based on calculation of the effective solvation forces acting on the biomolecule by using advanced solvation force extrapolation (ASFE) at inner time steps while solving the 3D-RISM integral equations only at large outer time steps.

The ASFE method is the most advanced implementation of RISM-MD up to date and it comprises a number of innovative theoretical and computational techniques such as: "non-Eckart transformation of coordinate space, modified least-square minimization of the residuals between the extrapolated and original coordinates/forces in the transformed space, maintaining an extended force-coordinate pair set while selecting the best subset for the smallest force residual minimum, balancing the normal equations, and incremental increase of the extrapolation interval."[27] In combination with the the optimized isokinetic Nosé-Hoover chain (OIN) thermostat,[26] this technique allows one to dramatically increase the accessible simulation time-scale for RISM-MD (up to several orders of magnitude).[27]

4.3   Examples of RISM implementation in molecular modelling software

Solvers for the 1D or 3D RISM equations have been implemented in several multi-purpose molecular modeling programs and stand-alone codes. The electronic structure of a solute in solution can be modeled using DFT coupled with the 3D RISM solvent model in the Amsterdam Density

Functional (ADF) program.[10] The Amber and AmberTools[1] codes permit 3D RISM to be used as a solvent model in conjunction with single-point, energy minimization, molecular dynamics and binding free energy calculations (MM-3DRISM) using standard force-fields. The output files of the Amber/AmberTools programs are in the correct format for input to the Placevent program,[193] which can be used to map solvent binding sites on proteins and other biological macromolecules. The CHARMM (Chemistry at HARvard Molecular Mechanics)[262] program contains an implementation of 1D RISM, but as yet there is no 3D RISM solver. The Molecular Operating Environment (MOE)[11] software uses an implementation of 3D RISM to map solvent and small-molecule binding sites on protein surface. The RISM-MOL[201] and RISM-MOL3D[4] packages contain routines to solve OZ type integral equations using a universal multigrid technique.[239,252,253] At the time of writing, the official release of NWCHEM[296] does not contain an implementation of 1D or 3D RISM methods, but a recent publication suggests that one or both of these methods may become available in the near future.[297]

Recently Maruyama et al developed a new highly efficient 3D-RISM code for massively parallel machines combined with the volumetric 3D fast Fourier transform (3D-FFT).[8] The code was tested on the RIKEN K supercomputer in Japan using a large calculation cell (2048^3 grid points) on 16,384 nodes, each having eight CPU cores; the results have shown that the new 3D-RISM has excellent scalability on the RIKEN K supercomputer.[8] This report opens new avenues for RISM applications on highly parallel supercomputer architectures.

5   Thermodynamic parameters within RISM

The total and direct correlation functions calculated by RISM allow one to obtain many different thermodynamic properties (partial molar volumes, solvation free energies, enthalpies, entropies, etc.) from analytical formulae. In this section we will overview main methods of calculations of these properties by RISM with a focus on recent developments in SFE/HFE calculations by RISM. At the end of this section we will provide a summary of main RISM SFE functionals developed up

to date in Table 1.

5.1 Partial molar volume (PMV) calculations

Partial molar volume (PMV) $\bar{V}$ is an important physicochemical property that characterizes how the volume of a solution varies with the addition of component i to the system at constant temperature and pressure:

$$\bar{V} = \left(\frac{\partial V}{\partial n_i} \right)_{T,P,n_{j \neq i}}.$$ (59)

PMV is a *thermodynamic* quantity and carries a different meaning from the geometric volume (i.e. molecular volume, van der Waals volume, etc.) of the solute.[192,300] At infinite dilution, the PMV of a molecular solute comprises the following terms:[192,299,301]

$$\bar{V} = \underbrace{V_W + V_T}_{V_C} + V_I + k_B T \chi_T,$$ (60)

where V_I ("interaction" volume) corresponds to a decreasing of the system volume because of specific interactions between water molecules and the solute molecule, V_C ("cavity" volume) is the volume of the cavity created in the solvent large enough to accommodate the solute molecule, $k_B T \chi_T$ is the ideal term that describes the volume effect related to the kinetic contribution to the
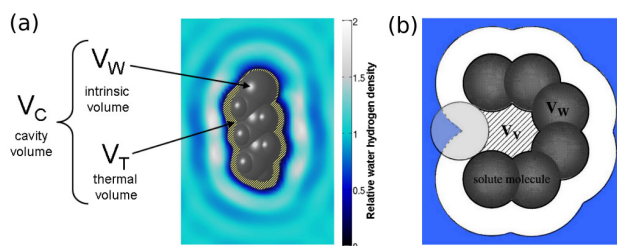
Figure 9: (a) The cavity volume of small solute can be represented as a combination of two contributions: "intrinsic" volume and "thermal" volume V_T ; (b) Schematic illustration of the van der Waals volume V_W and the void volume V_V within the solvent-inaccessible core of protein. Subfigure (a) is reproduced with permission from Ref.[298] Copyright 2011 American Institute of Physics. Subfigure (b) was adopted with permission from Ref.[299] Copyright 2001 John Wiley & Sons.

pressure of a solute molecule due to translational degrees of freedom, where χ_T is the isothermal compressibility of pure solvent at temperature T, k_B is the Boltzmann constant. For water, the value of the ideal term is about 1 $cm^3 \cdot mol^{-1}$,[301] and therefore it can usually be ignored. Chalikian et al[301] have shown that for proteins it is necessary to take into account the volume of structural voids within the solvent-inaccessible core of the solute molecule (V_V) (see Figure 9, b)).

The cavity volume, V_C can be further decomposed into two contributions: (i) the "intrinsic" volume of the solute (which can be approximated by the van der Waals volume V_W for low molecular weight solutes). (ii) the "thermal" ("void, "empty"[192,300]) volume V_T associated with thermally induced molecular vibrations of both the solute and solvent molecules, which create empty space around the solute molecule[299,301] (see Figure 9, a))

Dimensionless PMV (DPMV) $\rho\bar{V}$ in water can be calculated within the framework of the 1D RISM approach for the case of infinitely diluted solution using the following expression:[136,143]

$$\rho\bar{V} = 1 + \frac{4\pi\rho}{N}\sum_s\int_0^\infty \left(h_{oo}^{\text{solv}}(r) - h_{so}(r)\right)r^2 dr, \tag{61}$$

where N is the number of atoms in the solute molecule, and $h_{oo}^{solv}(r)$ and $h_{so}(r)$ are the total correlation functions between oxygen atoms in bulk water and between solute site s and water oxygen, respectively.

Within the 3D RISM approach, *solute-solvent site* correlation functions can be used to compute the solute DPMV via following expression:[188,192,299]

$$\rho\bar{V} = \rho k_B T\chi_T\left(1 - \rho\sum_{\alpha=1}^{N_{solvent}}\int_{R^3} c_\alpha(\mathbf{r})d\mathbf{r}\right). \tag{62}$$

PMV computed by both 1D RISM ($R = 0.984$) and 3D RISM ($R = 0.988$) are well correlated with experimental data for many different classes of chemicals[233,298] (see Figure 10). However, a slight bias that is linearly related to the solute size is observed for both 1D and 3D RISM. In Refs[233,298] a formula to correct the PMV calculations by RISM has been proposed, which takes
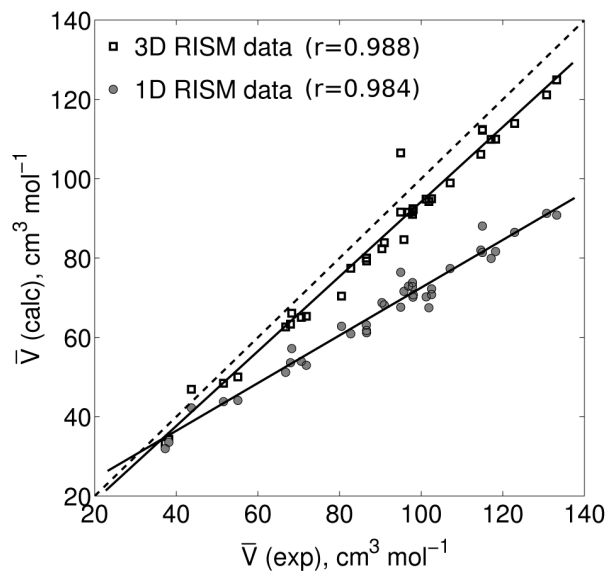
Figure 10: Comparison of experimental and calculated PMVs for a dataset comprising alcohols, alkanes, alkylbenzenes, dienes, ethers, ketones, and several polyfragment solutes. The ideal correlation is illustrated by the dashed line, while the line-of-best-fit is indicated by the solid line; r is the correlation coefficient. Republished with permission from Ref.[233] Copyright 2011.

into account the linear behavior of errors in both the 1D and 3D RISM approaches:

$$\bar{V} = b_1^{RISM} \cdot \bar{V}^{RISM} + b_0^{RISM}, \tag{63}$$

where $\bar{V}^{RISM}$ is the PMV obtained by the RISM approach, b_1^{RISM} is a scaling coefficient, b_0^{RISM} is an intercept.

Ratkova and Fedorov[298] demonstrated universal relations between the PMV in water and the static electric polarizability (α). They found that the relationship PMV $- \alpha$ has fine well-organized structure that strongly depends on the chemical nature of the solutes. Thus, for compounds within a specific chemical class the PMV $- \alpha$ correlation has a strong linear behavior. For different chemical classes these correlations are parallel-shifted with respect to each other. It was found that these shifts are caused by the specific solute-solvent interactions and can be presented as a function

of the solute lone pairs and π-electrons. This concept resulted in the following universal model: [298]

$$\bar{V} = A\alpha + B - (C_1 n_{lp} + C_2 n_{\pi e}),\tag{64}$$

where A, B, C_1 and C_2 are constants, n_{lp} is the number of lone pairs in a solute molecule and $n_{\pi e}$ is the number of π-electrons, accordingly. The efficiency and universality of the model was demonstrated on an external test set containing several dozen polyfunctional and druglike molecules.

PMV is an important property in interpreting the pressure induced denaturation of proteins. Imai et al[299] used 3D RISM and Kirkwood-Buff theory to calculate the PMVs of extended and α-helical forms of poly-glycine and poly-glutamic acid peptides containing from 1 to 7 residues (n.b. glutamic acid residues were modelled in the neutral form). For the transition from extended to helical form in poly-glycine oligomers, it was shown that $\Delta V_W \sim 0$, $\Delta V_V > 0$, $\Delta V_T < 0$, and $\Delta V_I \sim 0$, which indicates that for these peptides V_V and V_T are the most important terms in interpreting the conformational dependence of the total PMV, even though V_W typically makes an order of magnitude larger contribution to the absolute partial molar volume. The results were qualitatively similar for poly-glutamic acid, but the V_V term of the α-helical conformer was observed to contain contributions due to voids caused by the stacking of side-chains, as well as from the centre of the α-helix. Also, unlike glycine-oligomers, the interaction volumes (ΔV_I) of poly-glutamic acid peptides were observed to be uniformly negative, due to electrostriction of the solvent induced by the polar side chain groups. It was found that as the length of the peptide is increased, the Van der Waals volume (V_W) increases significantly more rapidly than the thermal volume. The corollary of this is that the PMV of a protein or similar large biomolecule is almost completely given by the sum of the Van der Waals and void volumes, both of which may be calculated from geometric considerations. This observation may explain why the use of molecular volumes rather than partial molar volumes in many empirical modelling studies has succeeded in accounting for volumetric properties of proteins.[302] It is interesting to note, however, that for small solutes (e.g. glycine or glutamic acid residue, or by inference a pharmaceutical) the thermal volume makes a significant

56

contribution to the total PMV. Imai et al[302] used 3D RISM to calculate the partial molar volumes of five proteins (bovine panreatic trypsin inhibitor, ribonuclease A, hen egg-white lysozyme, hovine milk β-lactoglobulin A, and bovine pancreatic α-chymotrypsinogen A) in pure aqueous solution at $25\,°C$. An excellent correlation was observed between the experimental and calculated partial molar volumes.

5.2 Enthalpy and entropy of solvation in 1D and 3D RISM

More detailed analysis of the solution-phase thermodynamics requires decomposition of the solvation free energy into the solvation enthalpy (ΔH_{solv}) and the solvation entropy (ΔS_{solv}). There are several ways to perform this decomposition.[168,191,208,303,304] The most general one is to determine entropy as a partial derivative of SFE over temperature.[191] Since experiments are most commonly performed at fixed pressure, it is convenient to introduce the decomposition as follows:

$$\Delta S_{solv} = \left(\frac{\partial \Delta G_{solv}}{\partial T} \right)_P ; \tag{65}$$

$$\Delta H_{solv} = \Delta G_{solv} + T \Delta S_{solv}.$$

This method requires calculation of the total and direct correlation functions at several values of the temperature.[191]

ΔH_{solv} contains two contributions from both solute-solvent (uv) and solvent-solvent (vv) interactions:[56,191]

$$\Delta H_{solv} = \Delta E_{uv} + \Delta E_{vv}, \tag{66}$$

where ΔE_{uv} is the energy of solute-solvent interactions and ΔE_{vv} is the solvent reorganization energy defined as the change in the solvent-solvent interaction energy due to the solute.[56]

The ΔE_{uv} term can be calculated relatively easily in both 1D and 3D-RISM formalisms[191,208,303] because it is a function of only the solute-solvent correlation functions. The resulting integrals are

given as:

$$\text{1D RISM: } \Delta E_{uv} = \rho \sum_{s\alpha} \int (h_{s\alpha}(r) + 1)u_{s\alpha}(r)dr, \tag{67}$$

$$\text{3D RISM: } \Delta E_{uv} = \rho \sum_{\alpha} \int (h_{\alpha}(\mathbf{r}) + 1)u_{\alpha}(\mathbf{r})d\mathbf{r}. $$

The term ΔE_{vv} depends on both temperature derivatives of distribution functions and thermodynamic ensemble.[191,305] Therefore, in practice it is most easily computed using the following expression:[?]

$$\Delta E_{vv} = \Delta G_{solv} + T\Delta S_{solv} - \Delta E_{uv}. \tag{68}$$

5.3 SFE functionals: large variety of expressions

Within the RISM approach, SFE is determined in terms of the pair distribution function, $g(\mathbf{r}_1, \mathbf{r}_2)$. Generally, the form of the SFE functional can be obtained by thermodynamic integration[306] using Kirkwood's equation:

$$\Delta G_{solv} = \int_0^1 d\lambda \left\langle \frac{\partial U(\mathbf{r}_1, \dots, \mathbf{r}_N, \lambda)}{\partial \lambda} \right\rangle_\lambda, \tag{69}$$

Within the 1D RISM approach, eq 69 can be written in the following form:

$$\Delta G_{solv}^{RISM} = \frac{\rho^2}{N} \sum_{s\alpha} \int_0^1 d\lambda \iint u_{s\alpha}(|\mathbf{r}_2 - \mathbf{r}_1|) \times$$
$$g_{s\alpha}(\mathbf{r}_1, \mathbf{r}_2, \lambda)d\mathbf{r}_1 d\mathbf{r}_2. \tag{70}$$

To evaluate this equation requires calculations of the total correlation function $h_{s\alpha}(r, \lambda)$ (see eq 33) at various λ. To determine the SFE for one compound one should perform about $10 - 100$ computer simulations, which in the case of complex organic molecules requires large computational resources.

5.3.1 End-point SFE functionals in 1D RISM

Singer and Chandler[204] showed that for the HNC closure relation the thermodynamic integration can be taken analytically and eq 70 can be replaced by a simpler SFE functional which allows one to obtain the value of ΔG_{solv} from a *single-point* computer simulation:

$$\Delta G_{solv}^{HNC} = 2\pi\rho k_B T \sum_{s\alpha} \int_0^\infty [-2c_{s\alpha}(r)-$$
$$c_{s\alpha}(r)h_{s\alpha}(r) + h_{s\alpha}^2(r)]r^2 dr. \tag{71}$$

Later Kovalenko and Hirata[143,307] performed a similar derivation of the 1D RISM/KH free energy functional:

$$\Delta G_{solv}^{KH} = 2\pi\rho k_B T \sum_{s=1}^N \sum_{\alpha=1}^M \int_0^\infty [-2c_{s\alpha}(r)-$$
$$c_{s\alpha}(r)h_{s\alpha}(r) + h_{s\alpha}^2(r)\Theta(-h_{s\alpha}(r))]r^2 dr. \tag{72}$$

However, it was shown that the energetic parameters of the system under investigation predicted by these free energy expressions are considerably overestimated.[134,140] Many efforts have been spent to improve the theoretical background of the RISM-based functionals for SFE calculations. Several advanced models have been developed to describe thermodynamics of hydration/solvation more accurately than previous methods. One of the earliest models developed by Chandler, Singh and Richardson assumes *Gaussian fluctuations* (1D RISM/GF) of the solvent molecules:[291]

$$\Delta G_{solv}^{GF} = 2\pi\rho k_B T \sum_{s=1}^N \sum_{\alpha=1}^M \int_0^\infty [-2c_{s\alpha}(r)-$$
$$c_{s\alpha}(r)h_{s\alpha}(r)]r^2 dr \tag{73}$$

Although the 1D RISM/GF free energy expression provides better agreement with experimental data for some solutes,[132] it is not widely used because it does not properly account for molecular effects for polar solutes.[135] More recently, Ten-no and Iwata[308] have proposed the *Partial Wave* (1D RISM/PW) model, which is derived from a distributed partial wave expansion of solvent molecules

around the solute:

$$\Delta G_{solv}^{PW} = \Delta G_{solv}^{GF} +$$
$$2\pi\rho k_B T \sum_{s=1}^{N} \sum_{\alpha=1}^{M} \int_0^\infty \tilde{h}_{s\alpha}(r) h_{s\alpha}(r) r^2 dr \tag{74}$$

where $r = |\mathbf{r_1} - \mathbf{r_2}|$ and

$$\tilde{h}_{s\alpha}(|\mathbf{r_2} - \mathbf{r_1}|) = \sum_{s'=1}^{N} \sum_{\xi=1}^{M} \int_{R^3} \int_{R^3} \tilde{\omega}_{ss'}(|\mathbf{r_1} - \mathbf{r'}|) \times$$
$$h_{s'\xi}(|\mathbf{r'} - \mathbf{r''}|) \tilde{\omega}_{\alpha\xi}^{solv}(|\mathbf{r''} - \mathbf{r_2}|) d\mathbf{r'} d\mathbf{r''},$$

$\tilde{\omega}_{ss'}(r)$ and $\tilde{\omega}_{\alpha\xi}^{solv}(r)$ are the elements of matrices $\mathbf{W}^{-1}$, $\mathbf{W}_{\text{solv}}^{-1}$ which are inverses to the matrices $\mathbf{W} = [\omega_{ss'}(r)]_{N_{\text{solute}} \times N_{\text{solute}}}$ and $\mathbf{W}_{\text{solv}} = [\omega_{\alpha\xi}^{solv}(r)]_{N_{\text{solvent}} \times N_{\text{solvent}}}$ built from the solute and solvent intramolecular correlation functions $\omega_{ss'}(r)$ and $\omega_{\alpha\xi}^{solv}(r)$ respectively.

Kovalenko and Hirata also proposed the repulsive bridge extension of the 1D RISM/HNC functional – the 1D RISM/HNCB free energy expression, to improve overestimated solute-solvent interactions for hydrophobic molecules:[208]

$$\Delta G_{solv}^{HNCB} = \Delta G_{solv}^{HNC} +$$
$$2\pi\rho k_B T \sum_{s\alpha} \int_0^\infty (h_{s\alpha}(r) + 1)(e^{-B_{s\alpha}^R(r)} - 1) r^2 dr. \tag{75}$$

Here $B_{s\alpha}^R(r)$ is a repulsive bridge correction function, which is given for each pair of solvent α and solute s atoms by the expression:

$$\exp(-B_{s\alpha}^R(r)) = \prod_{\xi \neq \alpha} \left\langle \omega_{\alpha\xi} * \exp\left(-\beta\varepsilon_{s\xi}\left(\frac{\sigma_{s\xi}}{r}\right)^{12}\right) \right\rangle \tag{76}$$

where $\omega_{\alpha\xi}(r)$ are the solvent intramolecular correlation functions, and $\sigma_{s\xi}$ and $\varepsilon_{s\xi}$ are the site-site parameters of the pair-wise Lennard-Jones potential.

Although reasonable qualitative agreement with experimental data have been reported by some

authors, benchmark studies [132,134–137] have shown that the standard RISM approaches do not deliver accurate SFEs; experimental and computed values may differ by an order of magnitude. So, despite the large variety of free energy functionals, accurate calculation of SFEs within RISM has been impossible until the recent development of novel semi-empirical free energy functionals (see the section Semi-empirical SFE functionals).

5.3.2 SFE functionals in 3D RISM

The 3D RISM/HNC free energy expression is given by:[309]

$$
\Delta G_{solv}^{HNC} = k_B T \sum_{\alpha=1}^{N_{solvent}} \rho_\alpha \int_{R^3} [\frac{1}{2}h_\alpha^2(\mathbf{r}) - c_\alpha(\mathbf{r})
$$
$$
- \frac{1}{2}c_\alpha(\mathbf{r})h_\alpha(\mathbf{r})]d\mathbf{r}. \tag{77}
$$

As shown by Kast and Kloss,[207] for a PSE-n expansion of the HNC closure of order n (see eq 47), corresponding free energy expression can be found analytically as:[207,209]

$$
\Delta G_{solv}^{PSE-n} = \Delta G_{solv}^{HNC}
$$
$$
- k_B T \sum_{\alpha=1}^{N_{solvent}} \rho_\alpha \int_{R^3} \left[\frac{\Theta\left(\Xi_\alpha\right)\right)\left(\Xi_\alpha(\mathbf{r})\right)^{n+1}}{(n+1)!} \right] d\mathbf{r}, \tag{78}
$$

where Θ is the Heaviside step function:

$$
\Theta(x) = \begin{cases} 1 & for \quad x > 0, \\ 0 & for \quad x \leq 0. \end{cases} \tag{79}
$$

As one can see, at $n \to \infty$ eq 78 corresponds to the 3D RISM/HNC free energy expression given by eq 77. Consequently, at n=1 eq 78 gives the 3D RISM/PLHNC (or 3D RISM/KH) free energy functional originally developed by Kovalenko and Hirata[309] that can be written in a simplified

form as:[3]

$$\Delta G_{solv}^{KH} = k_B T \sum_{\alpha=1}^{N_{solvent}} \rho_\alpha \int_{R^3} [\frac{1}{2} h_\alpha^2(\mathbf{r})\Theta(-h_\alpha(\mathbf{r}))$$
$$- c_\alpha(\mathbf{r}) - \frac{1}{2} c_\alpha(\mathbf{r}) h_\alpha(\mathbf{r})] d\mathbf{r}. \tag{80}$$

The GF free energy expression initially developed by Chandler, Singh and Richardson for 1D RISM[291] has been adopted by for the 3D RISM case:[101]

$$\Delta G_{solv}^{GF} = k_B T \sum_{\alpha=1}^{N_{solvent}} \rho_\alpha \int_{R^3} [-c_\alpha(\mathbf{r})-$$
$$\frac{1}{2} c_\alpha(\mathbf{r}) h_\alpha(\mathbf{r})] d\mathbf{r} \tag{81}$$

where ρ_α is the number density of a solute sites α.

Although the GF free energy expression and HNC-like free energy expressions (generally described by eq 78) have been extensively used to *qualitatively* model thermodynamics of different chemical systems by 3D RISM[3,101,192,304,310] they generally give HFE values that are strongly biased from experimental data with a large standard deviation error.[2,92,101,134,140,141] However, the advantage of these SFE functionals is that they all can be written in a form:

$$\Delta G_{solv}^{RISM} = \sum_{\alpha=1}^{N_{solvent}} \int_{R^3} \Phi_\alpha(\mathbf{r}) d\mathbf{r} \tag{82}$$

where the integrand functions $\Phi_\alpha(\mathbf{r})$ are given as compact analytical expressions. Therefore, the SFE can be decomposed into components coming from each of solute and solvent interaction sites that provide very useful information for studying detailed mechanisms or solute-solvent interactions and association of molecules in a complex solution.[7,194,195,311] The integrand functions $\Phi_\alpha(\mathbf{r})$ determine the 3D Solvation Free Energy Density (3D-SFED) coming from a given solvent

interaction site α around the solute molecule.[195] 3D-SFED characterizes the intensity of the effective solute-solvent interactions in different 3D spatial regions at the solute surface and indicates where these contribute the most/least to the entire SFE.[194] Therefore, 3D-SFED can provide useful information about different association/adsorption processes at complex biomolecule and supramolecule aggregate surfaces.[194,195,312]

5.4   Semi-empirical SFE functionals

Recently a new class of semi-empirical SFE functionals have been developed for both 1D and 3D RISM.[2,92,313] The unifying factor in these models is that they include some empirical corrections obtained by fitting against the error ε between calculated SFEs and experimental data:

$$\varepsilon = \Delta G_{solv}^{exp} - \Delta G_{solv}^{model}, \tag{83}$$

where ΔG_{solv}^{exp} is the experimental value of SFE, ΔG_{solv}^{model} is the SFE calculated by the RISM approach (superscript $model$ denotes the RISM-based free energy functionals, see the Section SFE functionals: large variety of expressions).

   Although a variety of different semi-empirical SFE functionals have been proposed for both 1D and 3D RISM,[2,92,136,141,202,313,315] they all use empirical parameters taken from one (or both) of the following classes: (i) a calculated partial molar volume; (ii) a set of atomic or substructure based descriptors. These classes of parameters are used to account for errors in the calculated SFEs that are systematic with respect to the excluded volume and selected chemical groups, respectively. In the following sections, we discuss the relative importance of each of these classes of descriptors for 1D and 3D RISM semi-empirical free energy functionals.

5.4.1   Improvement of SFE predictions with a volume correction

As discussed in Section 5.3, a wide variety of free energy functionals have been proposed for 1D (and 3D) RISM. However, several recent systematic studies[2,92,141] of the performance of common

RISM SFE functionals for predicting solvation thermodynamics in aqueous solutions have indicated that the accuracy of HFEs calculated by these methods is not satisfactory and for organic solutes the calculated values may differ by as much as an order of magnitude from experimental values. The errors in HFEs calculated by 1D RISM have previously been attributed to an overestimation of the energy required to form a cavity in the solvent and an underestimation of the electrostatic contribution to the hydration free energy of hydrogen bonding sites.[136] To address these deficits, Chuev et al[136] proposed the *Partial Wave Correction* (1D RISM/PWC) model:

$$\Delta G_{hyd}^{PWC} = \Delta G_{hyd}^{PW} + a\rho\beta^{-1}\bar{V} + b\delta_{\text{OH}}, \tag{84}$$

where ΔG_{hyd}^{PW} is the HFE value computed by the 1D RISM/PW functional (eq 74), $\bar{V}$ is the PMV of the solute (see eq 61), ρ is the number density of water, and δ_{OH} is the delta-function which equals 1 if OH-group is present in the solute molecule, otherwise it equals zero, a and b are the correction coefficients which are determined by linear regression.[136]

For a test set of 19 organic solutes studied in this work,[136] the 1D RISM/PWC free energy functional provided better agreement with experimental data than the original 1D RISM/PW model.[136,316,317] In its current state, however, the 1D RISM/PWC model is not suitable for many *polyfunctional* organic solutes due to its limited parameterization and it therefore requires further development.

For 1D RISM, only a weak correlation is observed between the calculated PMV and the error in calculated HFEs.[92] By contrast, the error in HFEs calculated using 3D RISM/GF model has a very strong linear correlation with the 3D RISM calculated partial molar volume (see Figure 11), with $R = 0.99$ for a dataset of 185 simple organic molecules). Based on this observation, Palmer et al[313] proposed the 3D RISM/UC (*Universal Correction*) free energy functional:[313]

$$\Delta G_{solv}^{UC} = \Delta G_{solv}^{GF} + a_1^{GF}(\rho\bar{V}) + a_0^{GF}, \tag{85}$$

where ΔG_{solv}^{GF} is the SFE obtained by the 3D RISM/GF free energy functional (eq 81), $\rho\bar{V}$ is the DPMV, a_1^{GF} is the scaling coefficient, a_0^{GF} is the intercept. The 3D RISM/UC model has been
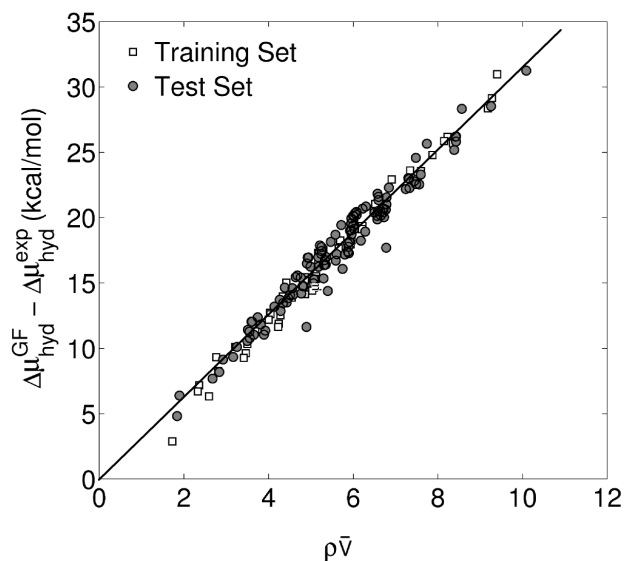
Figure 11: Correlation between the DPMV ($\rho\bar{V}$) and the difference between the 3D RISM calculated and experimental HFEs. Both the training and the test set molecules fall on the line-of-best-fit (which is indicated by a solid line). Reprinted with permission from Ref.[313] Copyright 2010 American Chemical Society.

shown to give accurate predictions of HFE for a test set of 120 organic compounds from different chemical classes (R=0.94, STDE=0.99 kcal/mol, ME=0.11 kcal/mol). The functional has been combined successfully with other molecular modelling methods to compute the intrinsic aqueous solubility of crystalline organic molecules[12] and to predict absolute and relative protein-ligand binding free energies in water.[102]

Recently, it was shown that the UC functional can be used for accurate prediction of octanol–water partition coefficient.[314] That work shown that the UC improves the prediction accuracy of 3D RISM/KH theory for SFE in octanol and reduces RMSE to as low as 1.03 kcal/mol. It was also noted in that work that "the UC makes the theoretical prediction less sensitive to a choice of solvent molecule geometry and possibly of force field parameters."[314]

The idea of a volume correction introduced in the 3D RISM/UC method[313] has been further developed in recently published works[9,315,318] which reported several new SFE functionals for molecular theories.

A *cavity-corrected* (CC) functional has been proposed which assumes that 3DRISM predicts

the electrostatic component of the HFE accurately, but requires a modification of the nonpolar contribution (i.e. the free energy for formation of the cavity created by the solute in the solvent) in order to yield accurate HFEs.[315] The method contains one adjustable parameter, γ, which can be computed from molecular dynamics simulations of Lennard-Jones solutes. The 3D RISM/CC functional has the form:

$$\Delta G_{hyd}^{cc} = \Delta G_{hyd}^{KH} + \frac{k_B T \rho_o}{2}(1 - \gamma) \int\limits_{V_{in}} c_o^{np}(\mathbf{r})d\mathbf{r}^3 \tag{86}$$

where ΔG_{hyd}^{KH} is the HFE computed by the KH free energy functional (eq 80), ρ_o is the number density of the water oxygen sites in bulk solution, and $c_o^{np}(\mathbf{r})$ is the direct correlation function for water oxygen sites computed with solute partial charges set to zero; $c_o^{np}(\mathbf{r})$ is integrated within the solvent excluded volume of the solute as indicated by V_{in}. The CC functional provides reasonably accurate estimates of the HFEs of small organic solutes ($R^2 = 0.88$ and RMSE=1.29),[315] but a validation of the method for druglike molecules has not yet been published.

5.4.2 Structural Descriptors Correction (SDC) functional: structural corrections as a tool to further improve SFE calculations

Basic ideas of the SDC model and its coupling with RISM (RISM/SDC model). The concept of using empirical corrections to improve the accuracy of SFEs calculated by RISM has recently been extended in *the Structural Descriptor Correction* (SDC) functional:[29,92]

$$\Delta G_{hyd}^{SDC} = \Delta G_{solv}^{model} + \sum_i a_i^{model} D_i + a_0^{model} \tag{87}$$

where ΔG_{sold}^{model} is the SFE calculated with a *model* free energy functional within the RISM approach, a_0^{model} is a constant, and the remaining terms are a set of structural corrections. The method is based on a combination of the RISM with a small number of chemical descriptors associated with the main features of the chemical structure of solutes. The main idea of the SDC model is illustrated by Figure 12.
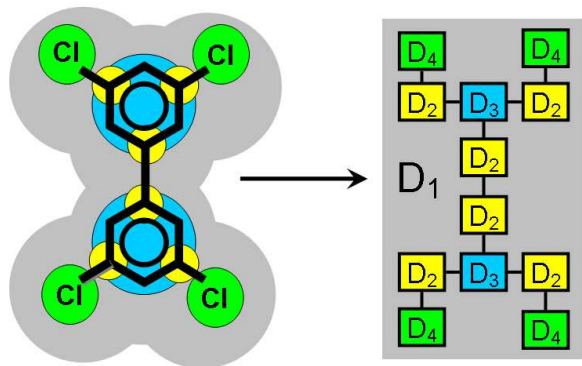
Structural Descriptors Correction (SDC) model

$$\Delta\mu_{hyd}^{SDC}=\Delta\mu_{hyd}^{RISM}+a_1\boxed{D_1}+a_2\boxed{D_2}+a_3\boxed{D_3}+a_4\boxed{D_4}+a_0$$

Figure 12: Illustration of the Structural Descriptor Corrections (SDCs) for one molecule from the dataset (3,3',5,5'-tetrachlorobiphenyl). The SDC model equation is a linear combination of the the following terms: $a_1 D_1$ is the DPMV correction, $a_2 D_2$ is the correction on branches, $a_3 D_3$ is the correction on benzene rings, $a_4 D_4$ is the correction on halogen atom, a_0 is the cavity independent systematic error (see eq 87). Reprinted with permission from Ref.[29] Copyright 2011 American Chemical Society.

The theoretical framework of the SDC model is as follows. Firstly, a small set of *structural* corrections (DPMV, counts of electron-donating/withdrawing substituents, counts of aromatic rings, etc) are introduced with empirical coefficients obtained by multi-linear regression so as to reduce the modelling error ε (see eq 83) of the RISM-based HFEs calculations. Since the structural correction descriptors are independent and the total number of the corresponding primary fragments is small, the regression can be carried out without concern over collinearity of the independent variables. However, the authors[92] selected only those variables that improved the accuracy of the model as quantified by the coefficient of determination, R^2, which is a commonly used metric to judge regression equations:[319]

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - f_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{88}$$

where f_i are the predicted values, n is the number of observations, y_i are the experimental values, $\bar{y}$ is the mean of the experimental data.
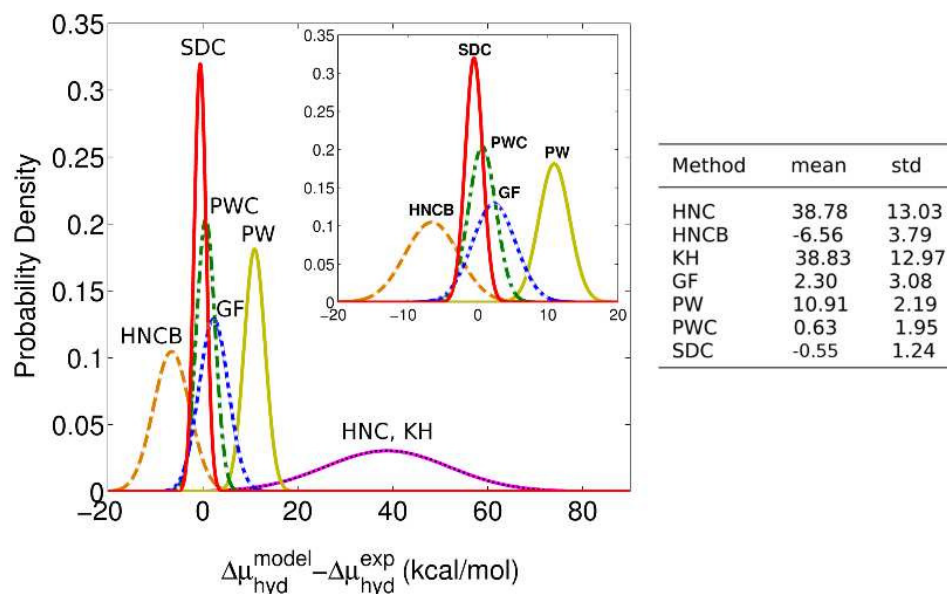
Figure 13: Normal distribution functions of the difference $\Delta G_{hyd}^{method} - \Delta G_{hyd}^{exp}$. The table on the right gives the accuracy of the HFEs calculated with different RISM HFE expressions in terms of the mean value (ME) and standard deviation (STDE) for a test set of 120 solutes (kcal/mol). The HNC closure was used in computing HFE with the HNC and HNCB functionals, while all other results were obtained with the PLHNC closure. The inset figure shows the same data for values between -20 kcal/mol and 20 kcal/mol on the x-axis. Reprinted with permission from Ref. [92] Copyright 2010 American Chemical Society.

Within the approach it is assumed that substructural elements of the solute contribute *independently* to the prediction error, which means that polyfunctional solutes can be represented as a linear combination of structural corrections and regression coefficients obtained for monofunctional solutes. Put another way, the SDC model should predict the property of interest for a wide range of polyfunctional solutes without additional reparametrization. For this reason, the training set was selected to contain only monofunctional molecules, whereas polyfunctional molecules were present only in the test set.

Significant structural descriptors. In the case of alkanes, that have no specific interactions with water molecules, the excluded volume effect makes a significant contribution to their hydration.[89,91,320] Several descriptors have been proposed which take into account the excluded volume effect: the solvent accessible surface area (SASA),[320–322] PMV,[136] their combination,[323] number of

68

carbon atoms,[324] etc. In previous works,[92,136] it has been shown that for a set of alkanes the error in HFE obtained by 1D RISM can be efficiently eliminated with a correction based on DPMV and a free coefficient a_0^{PW} (see eq 87, the last term). For solutes with branched alkyl chains, the number of branches can be incorporated as an additional correction to improve 1D RISM-calculated HFEs.[2]

The authors of these previous works[92,141] demonstrated that the empirical corrections described above are not enough to give accurate HFEs for other chemical classes of solutes. Each class of solutes has its own bias, but the standard deviation inside most of these classes is small. Thus, it was supposed that the bias for each chemical class can be removed by the use of the appropriate fragment correction. To obtain a higher accuracy of HFE predictions Ratkova et al[92] proposed a small set of fragment-based descriptors associated with the main features of the chemical structure of solutes: such as number of branches, electron-donating/withdrawing substituents, aromatic rings, etc. Alternatively, Palmer et al[141] used descriptors on number of atoms of different types in the molecules.

Performance of the 1D RISM/SDC model. Figure 13 shows the comparison analysis of the accuracy of HFEs calculated with different 1D RISM free energy functionals.[92] The 1D RISM/HNC and 1D RISM/KH free energy functionals (eqs 71, 72) give significantly overestimated values of HFE. The major component of the error is the bias of the difference $\Delta G_{hyd}^{model} - \Delta G_{hyd}^{exp}$ with respect to zero. However, removing the bias does not on its own lead to accurate HFE calculations for the 1D RISM/HNC and 1D RISM/KH free energy expressions because the STDE is considerable. By contrast, although HFEs calculated with the 1D RISM/PW free energy functional (eq 74) also contain a significant bias with respect to experimental data (ME is 10.91 ± 0.20 kcal/mol), they contain a much smaller random error. Thus, removing the bias of the error permits HFEs to be computed with high accuracy. Application of the 1D RISM/PWC model for the correction of 1D RISM/PW data leads to a decrease of the difference between experimental and calculated values (ME is 0.63 ± 0.18 kcal/mol) but it only weakly affects the STDE. In turn, the 1D RISM/PW data
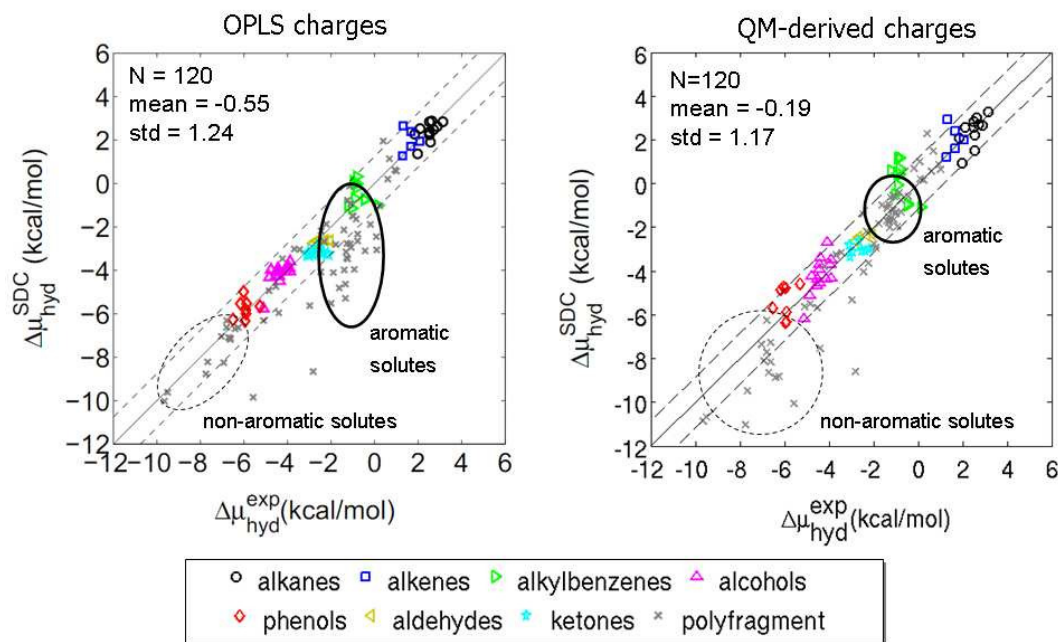
Figure 14: Performance of the 1D RISM/SDC model with different sets of partial charges for the internal test set of compounds. The ideal correlation is indicated with a solid line, while the STDE is given by dashed lines. Polyfragment aromatic solutes are indicated with solid circles, while dashed circles illustrate polyfragment non-aromatic solutes. Republished with permission from Ref.[233] Copyright 2011.

corrected with the 1D RISM/SDC model are less biased with respect to experimental values, while the STDE is significantly less than that for the PW method (about 1.5 times smaller). Thus, the HFEs computed using the SDC model are significantly more accurate than those obtained using the PW model.

It has been shown that the 1D RISM/SDC model allows one to obtain HFEs for "simple" solutes with high accuracy.[2,92] However, in the case of polyfragment solutes, the 1D RISM/SDC model is more sensitive to the chemical nature of solutes. Thus, the model allows one to predict HFEs with an accuracy of about 1 kcal/mol for chlorinated benzenes with fewer than three chlorine atoms but it provides worse results for chlorinated benzenes with larger number of chlorine atoms. Ratkova[233] have shown that reparameterizing the 1D RISM/SDC model using QM-derived partial charges (the 1D RISM/SDC(QMq) model) rather than OPLS partial charges leads to an increase in prediction accuracy. Since these calculations are significantly more computationally expensive,
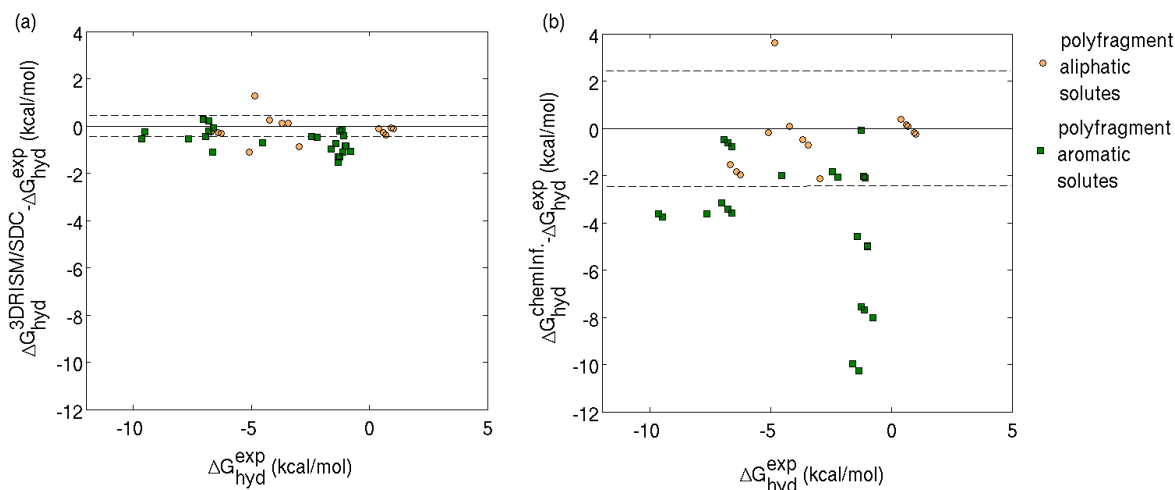
Figure 15: The difference between predicted and experimental HFEs versus experimental values for 38 polyfragment compounds of the test set of solutes: (a) the 3D RISM/SDC approach, (b) cheminformatics approach. Dashed lines indicate the corresponding standard deviations of the errors. Reprinted with permission from Ref.[2] Copyright 2011 American Chemical Society.

however, the authors recommend that for "simple" solutes the 1D RISM/SDC model should be used with OPLS partial charges, which are sufficient to give HFEs with small bias with respect to experimental data and a standard deviation of about 0.5 kcal/mol (see Figure 14). In the case of aromatic solutes, however, the QM-derived partial charges are more suitable since they are sensitive to the electron-donating/withdrawing nature of substituents, whereas the OPLS partial charges do not so accurately reflect these details.

Efficiency of the SDC functional with 1D and 3D RISM approaches. The accuracy of HFE calculations within the RISM approach can be further increased by employing the more advanced 3D RISM technique. The 3D RISM/SDC model requires fewer structural corrections compared to the 1D RISM/SDC model. In particular, since it estimates the cavity created by a solute molecule in water more accurately than 1D RISM, 3D RISM does not require a correction on number of branches in the solute.

The 1D RISM/SDC models and 3D RISM/SDC models have different but complementary attributes. Although the 3D RISM/SDC model is more accurate (RMSE ~ 0.5 kcal/mol rather than RMSE ~ 1.0 kcal/mol), it is also more computationally expensive (by ~ 100-fold). The authors of

71

these methods suggest that the 1D RISM/SDC model be used for initial high throughput screening of large compound databases, while the 3D RISM/SDC model is better suited to refinement of the screening results for selected compounds.

Comparison of the RISM/SDC with a cheminformatics approach. It has previously been demonstrated that the SDC functional yields more accurate HFE predictions with respect to traditional RISM-based free energy functionals.[2,92,202,313] Since the data obtained by the uncorrected free energy functionals (e.g. PW, GF, etc.) is biased, however, it is necessary to check whether its inclusion in the RISM/SDC models is warranted, or whether a cheminformatics model comprising regression against the SDC variables only would be more accurate. To investigate this further, the authors compared RISM/SDC and SDC only (i.e. omitting ΔG_{hyd}^{RISM} and DPMV ($\rho \bar{V}$)) models derived using the same fitting procedure and experimental data. For "simple" solutes, the cheminformatics based approach (SDC) gave similar results to the combined RISM and cheminformatics model (RISM/SDC model). However, very different results were obtained for polyfragment molecules, where HFEs computed by the cheminformatics (SDC) approach were found to be inaccurate, with prediction errors twice as large as those obtained with the RISM/SDC model (see Figure15). These results show that the RISM captures the most important features of the hydration phenomena that are not accessible in the cheminformatics approach.

5.5 Initial state correction SFE functionals

A conceptually different free energy functional has been recently proposed by Sergiievskyi et al that uses computed direct correlation functions to correct the SFE to the isothermal-isobaric ensemble.[318] The functional was initially developed for the Molecular Density Functional Theory (MDFT),[163,164,325–327] but can also be applied to 3D RISM.[318] The main idea behind this approach is to introduce a proper PMV correction for the correct initial state in the thermodynamic cycle of solvation to make the modelling results to be consistent with the isobaric-isotherm ensemble that is pertinent to experiments. Here we will refer to this SFE functional as the Thermodynamic

Ensemble Correction (TEC) functional. When applied to 3D RISM, the TEC functional takes the form:[9,318]

$$\Delta G_{solv}^{TEC} = \Delta G_{solv}^{RISM} + \frac{k_B T}{2}\rho^2 \hat{c}_s(k=0)\bar{V} \tag{89}$$

where ΔG_{solv}^{RISM} is the SFE computed from 3D RISM calculations, ρ is the number density of the pure solvent, $\hat{c}_s(k=0) \equiv \int c_s(|(r)|)d(r)$, and $\bar{V}$ is the PMV of the solute at infinite dilution in the given solvent. The second term in eq 89 can also be obtained from $\rho^2 \hat{c}_s(k=0) = 1 - 1/\rho k_B T \chi_T$. This approach has been tested by Sergiievskyi et al for both 3D RISM with the HNC closure and MDFT with the homogeneous reference fluid (HRF) approximation giving RMSDs of 1.8 kcal/mol and 2.39 kcal/mol, respectively, for a dataset of small organic molecules.[318] They obtained that for the 3D RISM/TEC the standard deviation of the error is considerably smaller than the RMSD ($\sigma_{3DRISM} = 1.45$ kcal/mol and $\sigma_{MDFT} = 1.6$ kcal/mol), which indicates a large systematic error that the authors attribute to the use of the HNC closure and other approximations behind the 3D RISM method.

The idea of the initial state correction suggested by Sergiievskyi et al[318] was further developed in the recent work by Misin et al.[9] The authors of that work suggested that for the 3D RISM theory another functional has to be used instead of eq 89 to make a correction for the proper initial state in the modeling thermodynamic cycle of solvation. The formula for this functional (referred as the Initial State Correction (ISc) functional) reads as:[9]

$$\Delta G_{solv}^{ISc} = \Delta G_{solv}^{RISM} - \rho k_B T \bar{V} + \frac{k_B T}{2}\rho^2 \hat{c}_s(k=0)\bar{V}. \tag{90}$$

To support their statement, Misin et al presented results of an extensive benchmarking of both the TEC and the ISc SFE functionals for 3D RISM calculations of HFEs of various organic molecules at a range of different temperatures between 278 K and 378 K.[9] The original derivation by Sergiievskyi et al of general formulae behind the thermodynamic ensemble/initial state correction was developed within the scope of the HNC approximation.[318] Therefore, Misin et al used

either the HNC SFE functional (eq 77) or the 3d order PSE-n expansion (PSE-3) of the HNC SFE functional (eq 78) for calculating the ΔG_{solv}^{RISM} term in both TEC and ISc SFE functionals.

The results of the benchmarking have shown that the RISM/ISc SFE functional provides better accuracy of the HFE predictions than the RISM/TEC SFE functional for the whole range of temperatures. In addition, the RMSE of the RISM/ISc HFE predictions remains practically constant for all temperature (around 1.5-1.7 kcal/mol). To compare, the RMSE of the RISM/TEC HFE predictions gradually increases with temperature from 3 kcal/mol at 278 K to almost 4 kcal/mol at 368 K.[9]

The 3D RISM/ISc model provides best accuracy when combined with the HNC closure/SFE functional (eq 46 and eq 77) for calculating the ΔG_{solv}^{RISM} term in eq 90; however, much better numerical stability with only a slight decrease in accuracy can be achieved with use of the PSE-3 closure/SFE functional (eq 47 and eq 78).[9]

We note that in contrast to the 3D RISM/UC (eq 85), 3D RISM/CC (eq 86) and 3D RISM/SDC (eq 87) SFE functionals that use adjustable parameters, 3D RISM SFE calculations by the TEC and ISc SFE functionals do not require additional training and/or extensive parameterization. At the same time, the 3D RISM/ISc model predicts the HFE for a wide set of organic molecules both at ambient and non-ambient temperatures with an accuracy that is comparable[9] with the accuracy of the HFE predictions provided by MD calculations for a similar set of molecules.[114]

The theoretical background behind the thermodynamic ensemble/initial state corrections is currently a subject of debates.[328] Nevertheless, the pilot results presented by Sergiievskyi et al[318] and Misin et al[9] look encouraging and give a hope that the accuracy of solvation thermodynamic parameters calculated by molecular theories can be significantly improved.

Table 1: SFE functionals within RISM approach

Free energy functional	Theory	Citation	Closure		Corrections	Comments
			correspon-ding	employed		
Hypernetted chain (HNC)	1D, 3D RISM	Singer et al 1985;[204] Kovalenko et al 1999[150]	HNC	HNC[92,134,135,140] KH[92]	no	The HNC and KH functionals were obtained by taking the thermodynamic integration analytically.[150,204] The both functionals provide significantly over-estimated values of HFE.[2,92,134,140,202,313] Sometimes the KH SFE functional is also called as PLHNC SFE functional.
Kovalenko-Hirata (KH)	1D, 3D RISM	Kovalenko et al 2000[307,309]	KH	KH[2,92,140]	no	

Table 1 – continued from previous page

Free energy functional	Theory	Citation	Closure		Corrections	Comments
			correspon-ding	employed		
PSE-n	3D RISM	Kast&Kloss 2008[207] (see also Ref.[209])	PSE-n	PSE-n [9,196,209]	series expansion of 3D HNC	The PSE-n SFE functional is a series expansion of the the 3D HNC SFE functional (eq 78) that corresponds to the 3D PSE-n closure (eq 47) that provides a better numerical convergency than the 3D HNC closure.[207,209] At $n \to \infty$ the 3D RISM/PSE-n SFE functional corresponds to the 3D RISM/HNC SFE functional; consequently, at n=1 it corresponds to the 3D RISM/KH SFE functional.[207]

Table 1 – continued from previous page

Free energy functional	Theory	Citation	Closure		Corrections	Comments
			corresponding	employed		
Hypernetted chain with repulsive bridge correction (HNC+RBC or HNCB)	1D, 3D RISM	Kovalenko et al 2000[208]	HNC	HNC[137,208] KH[92,137]	Repulsive bridge correction, eq 76.	The correction was proposed to improve overestimated solute-solvent interactions for hydrophobic molecules.
Gaussian fluctuations (GF)	1D, 3D RISM	Chandler et al 1984;[291] Genheden et al 2010[101]	–	HNC[134,135,140] KH[2,92,313,329]	no	The GF functional supposes Gaussian fluctuations of solvent. The drawback of the functional is the improper account of molecular effects for polar solutes.[92]

Continued on next page

Table 1 – continued from previous page

Free energy functional	Theory	Citation	Closure		Corrections	Comments
			correspon- ding	employed		
Partial wave (PW)	1D RISM	Ten-no et al 1999[308]	–	HNC[92,134,135,140] KH[2,29,92,136]	no	The PW functional is based on the dis- tributed partial wave expansion and more accurately accounts for angular correla- tions of solvent molecules around the so- lute. It sufficiently overestimates HFE for nonpolar solutes.[92]

Table 1 – continued from previous page

Free energy functional	Theory	Citation	Closure		Corrections	Comments
			correspon-ding	employed		
Partial wave correction (PWC)	1D RISM	Chuev et al 2007[136]	–	KH[92,136,316]	Two empirical corrections on the DPMV $(\rho\bar{V})$ and the presence of the OH-group, eq 84.	Values of the empirical coefficients a and b (eq 84) are obtained by linear regression against the experimental HFE values. The PWC functional predicts HFE more accurately than the original PW model. However, the functional is not suitable for many organic solutes due to its poor parameterization and requires further development.

Continued on next page

Table 1 – continued from previous page

Free energy functional	Theory	Citation	Closure		Corrections	Comments
			correspon-ding	employed		
Universal correction (UC)	3D RISM	Palmer et al 2010[313]	–	KH[2,313,329]	The empirical correction on DPMV, eq 85.	Values of the empirical coefficients a_1 and a_0 in eq 85 are obtained by linear regression against the training data. The predictive ability of the functional is assessed using coefficients obtained from the training data. The coefficients can be transferred between multiple different classes of organic molecules. The UC functional can be further improved using the larger set of structural corrections (see the SDC functional, eq 87).

Table 1 – continued from previous page

Free energy functional	Theory	Citation	Closure		Corrections	Comments
			correspon-ding	employed		
Structural descriptors correction (SDC)	1D, 3D RISM	Ratkova et al 2010,[92] Frolov et al 2011[2]	–	KH[2,29,92]	The set of structural empirical corrections derived from the solute molecule (DPMV, counts of aromatic rings, counts of electron-donating/withdrawing substituents, etc.), eq 87.	The empirical coefficients $\{a_i\}$ in eq 87 are evaluated by linear regression against the training data. The predictive ability of the functional is analyzed using the same set of coefficients as for the training set. The coefficients can be transferred between different classes of solutes. The SDC functional yields one of the most accurate HFE predictions with respect to other RISM-based HFE expressions.

Table 1 – continued from previous page

Free energy functional	Theory	Citation	Closure		Corrections	Comments
			corresponding	employed		
Thermodynamic Ensemble Correction (TEC)	3D RISM, MDFT	Sergiievskyi et al 2014[318]	HNC	HNC[9,318] PSE-3[9]	Thermodynamic Ensemble Correction, eq 89.	A new SFE functional has been recently proposed by Sergiievskyi et al that uses computed direct correlation functions to correct the SFE to the isothermal-isobaric ensemble.[318] The functional was initially developed for MDFT,[318] but can also be applied to 3D RISM.[9,318] The main idea behind this approach is to introduce a proper PMV correction for the correct initial state in the thermodynamic cycle of solvation.[318]

Continued on next page

Table 1 – continued from previous page

Free energy functional	Theory	Citation	Closure		Corrections	Comments
			correspon-ding	employed		
Initial State Correction (ISc)	3D RISM	Misin et al 2015[9]	HNC, PSE-3	HNC, PSE-3[9]	Initial State Correction, eq 90.	Later on Misin et al[9] shown that the ISc SFE functional (eq 90) provides more accurate results for SFE predictions for organic molecules by 3D RISM than the TEC functional (eq 89); the ISc method can be applied both at ambient and non-ambient temperatures with practically the same RMSE of around 1.5–1.7 kcal/mol.[9] 3D RISM SFE calculations by the TEC and ISc SFE functionals do not require additional training and/or extensive parameterization.

6 RISM coming into laboratories: selected examples of recent applications

Thanks to the recent advances in theory and computational discussed above, RISM has recently begun to be used more widely in the computational molecular sciences.

The main areas of application of RISM fall into three general categories:

- molecular-scale effects on solvent structure and thermodynamics with a particular focus on solvation phenomena at complex molecular interfaces (e.g. mapping of solvent and cosolute binding sites on the surface of biological macromolecules);

- solvent effects on conformational and configurational properties of bio- and macromolecules molecules and their assemblies;

- analysing and predicting phys-chem properties of molecular solvation.

In this section we provide several examples of successful RISM applications in different areas of life sciences, chemical engineering and biotechnology. More examples can be found in recent topical reviews and future articles.[3,6,7,15,138]

6.1 Solvation phenomena at complex molecular interfaces

6.1.1 Mapping of solvent and cosolute binding sites on the surface of biological macromolecules

1D and 3D RISM permit the equilibrium density distributions of solvent and co-solute species around a solute to be calculated at considerably lower computational expense than by molecular simulation. The obtained functions are free from statistical noise and by definition correspond to the thermodynamic limit. Since the maximums in these functions correspond to regions where there is a high density of the given solvent component, they provide an excellent method to map solvent and cosolute binding sites on solutes.
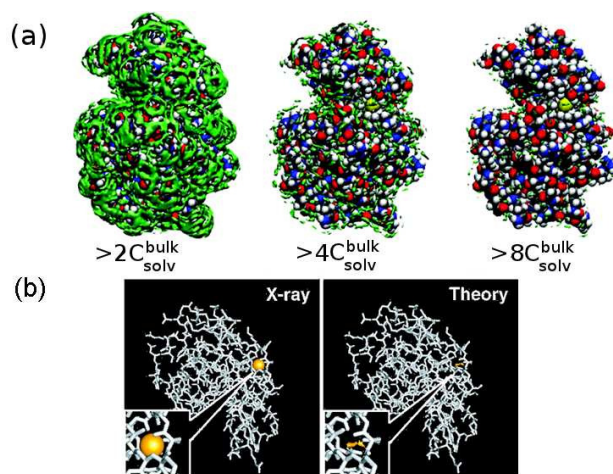
Figure 16: Mapping of solvent and co-solute binding sites on the surface of human lysozyme: (a) 3D map of high water density ($¿$ 2 times (left), 4 times (center), and 8 times (right) the bulk solvent density) at the surface of the protein calculated by the 3D RISM theory; (b) the left picture shows the calcium binding site in the protein detected by X-ray, while the picture in the right exhibits the binding site found by the 3D RISM theory. Reprinted with permission from Ref.[290] Copyright 2009 American Chemical Society.

3D RISM is particularly useful here because it allows one to calculate three dimensional solvent/cosolute density maps around highly complicated biomacromolecular surfaces. Moreover, the probable binding positions of the solvent/cosolute molecules can be reconstructed from the high solvent/cosolute density areas of the density correlation functions (Figure 16).

It has been shown that to an acceptable resolution the solvent distribution functions calculated by 3D RISM around a model protein are very similar to those obtained by MD using common solvent models (e.g. TIP3P, TIP4Pew),[22] which suggests that 3D RISM is a reliable method for mapping of solvent/cosolute binding in these systems. Moreover, 1D and 3D RISM can be used to model solvent components at a wide range of concentrations (from below mM to the saturation limit). The ability to model solutions with low concentrations of cosolute components is of particular importance for simulating solutions of poorly soluble drugs and pH effects on biomolecules since these low concentration regimes are difficult to sample adequately using explicit solvent simulations (e.g. MC or MD).

The prediction of structural and thermodynamic properties of the binding of small ligands (e.g.

water, ions, pharmaceuticals) by proteins is one of the most important challenges in computational chemistry. The "docking" simulations commonly used in pharmaceutical research use essentially a trial and error scheme to find the "best-fit complex" between the host and guest molecules based on geometric and/or energetic criteria. A significant problem with these approaches is the lack of solvent effects accounting. Particularly, they do not consider properly the desolvation penalty associated with displacing water molecules at the interface between host and guest molecule. Moreover, during the last decades, it was revealed that treatment of water molecules in binding cavity must be routinely included in "docking" studies since they can regulate the binding event via mediating of protein-ligand interactions (so-called "structural" water).[330–332] Exclusion of these water molecules from "docking" simulations leads to an improper description of binding, which is one of the crucial reasons for high rates of false-positive drug candidate at the early-stage drug design. Although there is a recent progress in the field,[333] accurate accounting of solvation effects (including water-mediated protein-ligand interactions, the competition between different solvent and ligand species for binding sites, etc.) remains one of topical challenges in "docking" studies.

The application of 3D RISM to molecular recognition problems including protein-ligand binding has significant potential. As previously mentioned, solution of the 3D RISM equations for a host molecule (protein) in a solvent containing the guest molecules (ligand) will in principle produce both the most probable binding mode and solvation thermodynamics including all enthalpy, entropy and free energy terms, and their derivatives such as compressibilities and partial molar volumes. In practice, this is a relatively new field that requires more research, but there have already been some significant successes reported. Imai et al[17] used the 3D RISM approach to successfully predict water binding sites in hen egg-white lysozyme. Although the study was limited in scope to only one protein in pure aqueous solvent, the results are significant because they show that the preferred binding sites of a guest molecule (in this case, water) to a host (protein) can be accurately predicted using spatial distribution functions calculated by 3D RISM.[290] This method was later implemented in the Placevent algorithm by Sindhikara et al[193] and independently validated for the prediction of water binding sites on the surface of bovine chymosin by Palmer et al.[102] Yoshida et

al[18,224] used a similar methodology to successfully find the Ca^{2+} binding sites in human lysozyme. In this approach, ions are included in the molecular model of the solvent. The ion binding sites are then identifiable as regions of high density in the spatial distribution functions of the ion (solvent) around the solute (protein), which are obtained by solving the 3D RISM equations. Kiyota et al[334] studied the binding of O_2, Xe, NO, CO, H_2S and H_2O by myoglobin using 3D RISM. Interestingly, those ligands that have a toxic effect on the body were observed to have higher binding affinities than oxygen. Imai et al[14] used a fragment-based approach coupled with 3D RISM to perform ligand-mapping calculations (3D-RISM-LM method) to probe the mechanism of the multidrug efflux transporter AcrB. Sixteen organic solvent molecules representing aromatic, aliphatic (hydrophobic) hydrogen-bond donor, and hydrogen-bond acceptor groups were used as drug fragments. Based on the most probable binding modes of these rigid fragments, the authors propose a mechanism for drug transport by the membrane protein. Similarly, Phongphanphanee et al[335] used 3D RISM to calculate the spatial distribution functions and potentials of mean force characterizing the transport of seven ligands (H_2O, Ne, CO_2, NO, NH_3, urea and glycerol) by two aquaporin channels, AQP1 and GlpF. The authors report that AQP1 transports Ne, H_2O and NH_3, while GlpF transports all seven ligands. Comparison of the results of Phongphanphanee, Yoshida and Hirata (from 3D RISM) with those of Hub and de Groot (from MD simulations) has prompted discussion about the formal definition of potentials of mean force for these systems.[336,337] Howard et al.[19] used a renormalized 3D RISM approach to compute both the water and ion spatial distributions around B-form DNA duplexes consisting of the $dA_n - dT_n$ and $dG_n - dC_n$ nucleotide sequences for $n = 1-12$. The calculations satisfactorily reproduced the results of explicit solvent simulations for 1M KCl, 0.1M KCl and 0.1M NaCl electrolyte solutions. As would be expected, the charged phosphate backbone of the DNA duplex was observed to be stabilised by a significant density of cations in all simulations, with Na+ ions predicted to bind more strongly than K+ ions. One step further in the direction of IET application for the rational drug design was done by Nikolic et al.[197] They coupled 3D RISM with the open-source AutoDock program (3D-RISM-DOCK), what allows treatment of flexible ligand using the following steps: the fragmental decomposition of a ligand,

mapping of its small rigid fragments, and reconstruction of the ligand molecule in its the most probable conformation in the binding pocket. The empirical AutoDock scoring function was completely replaced by PMF calculated by 3D RISM. Therefore, the ranking of docked conformations was performed using the free energy calculated within the same statistical mechanical framework and the same force field as used for "docking".

6.1.2 Molecule-surface recognition and supramolecular interactions

Recently 3D RISM/KH theory in combination with experimental techniques was used to study molecular recognition interactions in non-aqueous solvents between kaolinite and a series of heterocyclic aromatic compounds (HAC) that were chosen as representatives of different moieties in petroleum asphaltene macromolecules.[172,312] In these studies the 3D RISM/KH method was used to predict the adsorption configuration and thermodynamics from the 3D site density distribution functions and total SFE, respectively, for adsorption of HAC and solvents (toluene and heptane) on kaolinite. The results were used to interpret spectrophotometric measurements and experimentally determined adsorption isotherms. This combined experimental and computational modeling approach provided useful insights into the specific interactions among clays, bitumen, and solvents that can be used for development of environmentally friendly enhanced oil recovery (EOR) techniques for tight oil formations like e.g. oil sands in Canada.[172,312]

The 3D RISM/KH method was also successfully applied to study supramolecular interactions in Plant Cell Walls (PCWs).[194,195] The 3D-SFED maps calculated by 3D RISM were used to understand the molecular basis of the recalcitrance forces within the PCW. These studies investigated the effect of hemicellulose composition on nanoscale forces that control the primary PCW strength and nanoarchitecture[194] and mechanisms of lignin-lignin and lignin-hemicellulose thermodynamic interactions.[195] These works provided a detailed molecular view of the fundamental interactions within the PCWs that lead to recalcitrance. The results are important because they provide scientific background for the development of new technologies for plant biomass conversion (e.g. sustainable production of biofuel and chemical feedstock[338,339]).
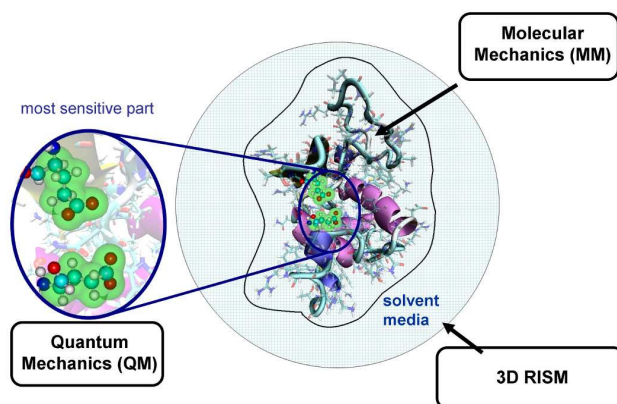
Figure 17: Schematic representation of the RISM–QM/MM method, where each part of the system is treated by the most suitable method.

6.2 Solvent effects on conformational and configurational properties of biomolecules and their assemblies

Solvent plays a significant role in mediating the structural dynamics of biological macromolecules and their assemblies. A variety of approaches based on 1D and 3D RISM have been used to study different solvent and co-solute effects on conformational and configurational transitions in biomolecules. The earliest of these studies used single point calculations of solvation free energies and solvent structure to determine the most stable members of preselected sets of molecular conformations/configurations. More recently, however, dynamic systems have been studied using 1D and 3D RISM coupled with geometry optimization, normal mode analysis, MD and MC simulations, to provide a more realistic description of their dynamics. Furthermore, RISM has recently been employed to develop multi-scale quantum-mechanics/molecular mechanics (QM/MM) methods to study biological macromolecules.[7,340,341] In this approach, a selected region of the solute (e.g. an enzyme active site) is treated by quantum mechanics, the surrounding region of the biomolecule is modeled at the molecular mechanics level, and the solvent effects are incorporated by RISM (see Figure 17). Such multiscale methods are excellent approaches to model quantum-level effects of biological macromolecules that are too large to treat with a pure quantum mechanics approaches.

Kinoshita et al.[177] coupled Monte Carlo simulated annealing with 1D RISM to study the pre-

ferred conformations of the Met-enkephalin peptide (Tyr-Gly-Gly-Phe-Met) in different solvents. It was shown that the conformation space of Met-enkephalin peptide could be sampled in fewer steps in aqueous solvent than in simple non-aqueous solvents because in the former many conformers were found to be too high in energy to be observed at biological temperatures. The authors speculate that this focusing of peptide conformations in water might go some way to address the paradox related to protein folding noted by Levinthal, who pointed out that the time for a random search of all possible conformations would be unrealistically long even for a relatively small peptide/protein. Similarly, Kinoshita et al[185] used 1D RISM to study the influence of water, methanol and ethanol solvents on four selected conformations of Met-enkephalin peptide (Tyr-Gly-Gly-Phe-Met) and two selected conformations of a peptide from ribonuclease A (Lys-Glu-Thr-Ala-Ala-Ala-Lys-Phe-Leu-Arg-Gln-His-Met). The total solvation free energy was shown to vary less between different peptide conformations in alcohols than in water, which was rationalised by partitioning the solvation free energy calculated by 1D RISM into contributions from individual solvophilic and solvophobic atoms. In alcohol, solvophobic atoms were shown to be less solvophobic than in water, and solvophilic atoms were found to be less solvophilic. Consequently, in alcohol where the solvation free energy varies less between different conformers, the peptides tend to adopt conformers that have the lowest internal energies (i.e. α-helix for Met-enkephalin peptide and the C-peptide tested in this paper, but alcohol is also observed to stabilise β-sheet structures in some peptides).

Kitao et al.[342] studied the effects of (pure) aqueous solvent on the conformation and collective motions of the small protein melittin (26 amino acids and 436 atoms) using both explicit solvent molecular dynamics and XRISM theory. The conformational energy profiles along the two lowest frequency normal modes of the protein were calculated as $\Delta F = \Delta G_{XRISM} + \Delta E_{conf}$, where ΔG_{XRISM} is the hydration free energy calculated by XRISM and ΔE_{conf} is the internal conformational energy of the protein. Maruyama et al.[343] used geometry optimization coupled with 3D RISM to study the preferred conformations of human telomere in different electrolyte solutions (pure water, 0.1M NaCl, 0.1M KCl). In agreement with experiment, the calculations predict that

the preferred conformation of the telomere is different in 0.1M NaCl ("basket" conformation) than in pure water ("chair" conformation). Yamazaki et al.[344] studied the self-assembly and stability of organic rosette nanotubes (RNT) from monomers comprised of guanine - cytosine blocks decorated with 4-aminobenzo-18-crown-6-ether. A putative self-assembly process was proposed and shown by 3D RISM calculations to be thermodynamically favorable. The salt-induced conformational change of the DNA oligomer d(5'CGCGCGCGCGCGCG3') from B-form to Z-form has been rationalized by Maruyama et al[345] on the basis of 3D RISM calculations. The conformational stability of the DNA oligomers was interpreted on the basis of Coulombic interactions between neighboring phosphate groups and counter-ions. In low salt concentrations, Coulombic repulsions between phosphate groups mean that the B-form is more favorable than the Z-form, primarily because the latter has a shorter interphosphate distance. In high salt concentrations, however, the counter-ions effectively screen the phosphate groups reversing the relative stabilities of the two forms. The results do not support the hypothesis of Saenger at al.[346]

The essential elements of life (self-replication, information processing, metabolism) occur largely by specific interaction between biological molecules. Understanding how two molecules recognize each other in biological environments is thus one of the fundamental issues in the biomolecular sciences. It is also of great importance in many areas of industrial research including the pharmaceutical industry, where predictions of the binding modes of small molecules (e.g. waters, ions, drugs) to biological macromolecules (e.g. proteins, DNA) are used to help design new pharmaceuticals. The key thermodynamic parameter characterizing the binding of a ligand (L) by a receptor (R) is the binding free energy (ΔG_{bind}) for the process $R + L \rightarrow RL$.

Since the solvent phase plays an important role in mediating many molecular recognition events in biological systems, it is vital that computational methods to model biomolecular recognition incorporate an adequate molecular-scale model of solvent. Several methods to compute absolute or relative binding free energies using molecular integral equation theory have been proposed. Genheden et al. computed the binding free energies of seven biotin analogues to avidin using both MM-PBSA and MM-3DRISM methods.[101] Blinov et al. used the MM-3DRISM method to
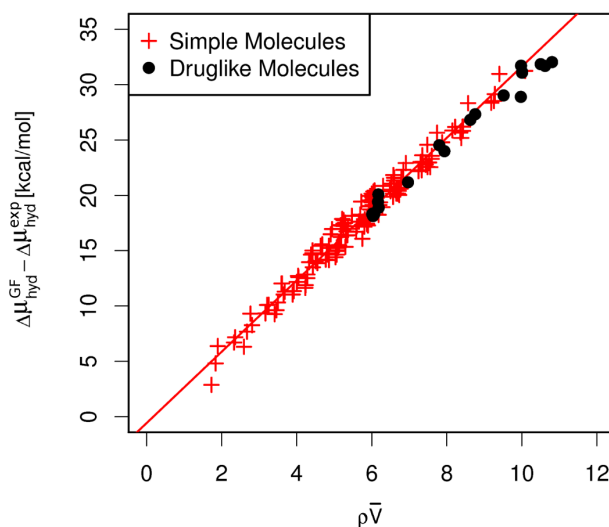
Figure 18: Correlation between the error in the HFEs calculated by the 3D RISM with the PLHNC closure using the GF free energy expression: $\Delta G_{hyd}^{GF} - \Delta G_{hyd}^{exp}$, and the DPMV, $\rho\bar{V}$. Red crosses are small organic molecules; black dots are druglike solutes. The pharmaceutical molecules lie on the line-of-best-fit calculated for simple organic molecules (the solid red line). Reprinted with permission from Ref.[329] Copyright 2011 American Chemical Society.

compute the binding thermodynamics of amyloid fibrils.[310] The association free energy of small β-sheet oligomers was observed to increase approximately linearly with the number of peptides in the complex. Palmer et al. performed computational alanine scanning calculations of the bovine chymosin – bovine κ-casein complex using MM-3DRISM to probe the contribution of individual residues to the overall binding free energy.[102] The complex is of industrial interest because bovine chymosin is widely-used to cleave bovine κ-casein and to initiate milk clotting in the manufacturing of processed dairy products.[347–349]

6.3 Analysing and predicting phys-chem properties of molecular solvation

Palmer et al.[329] have recently demonstrated that the empirical coefficients in the 3D RISM/UC are transferable to chemical classes other than those represented in the training data. For druglike molecules, it was observed that uncorrected HFEs computed by the 3D RISM/GF method also have a strong linear correlation with DPMV ($R = 0.99$). Indeed, the druglike molecules lie on the line-of-best-fit calculated for the simple organic molecules (see Figure 18). For 21 druglike molecules,
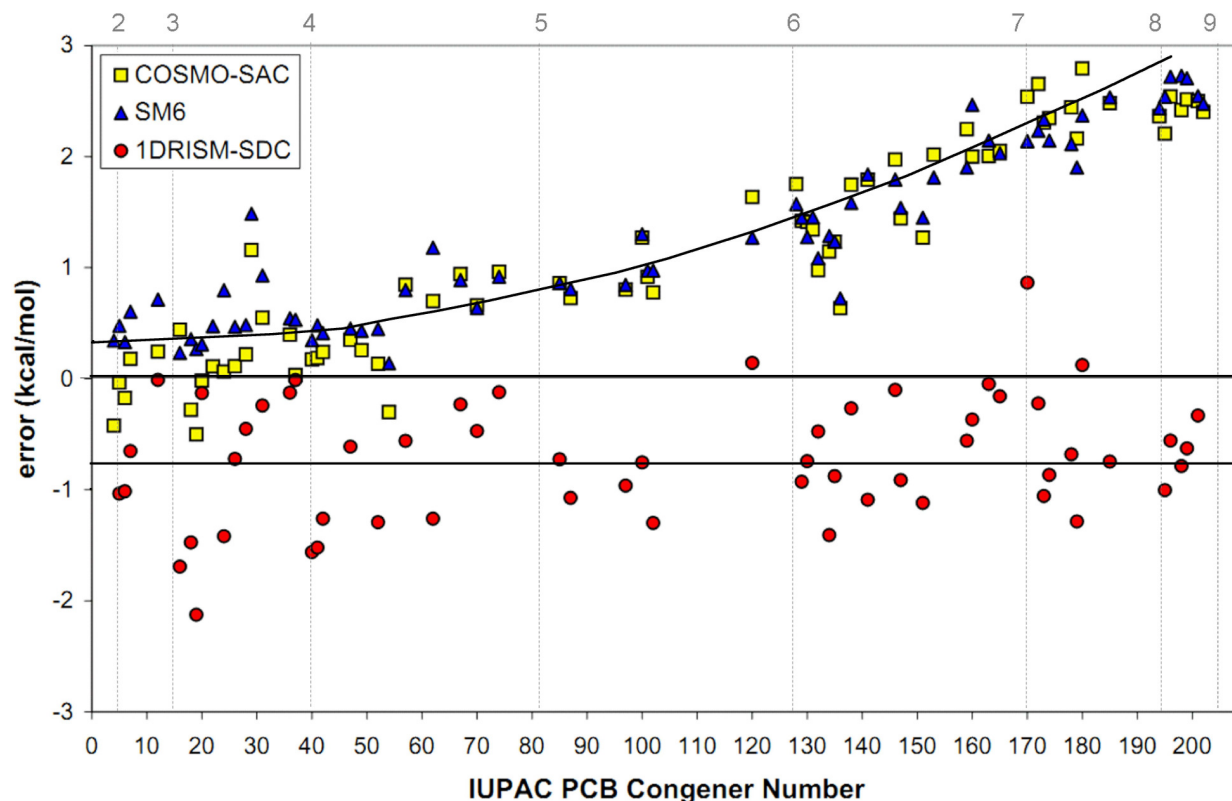
Figure 19: Error in HFE prediction for the 1D RISM/SDC(QMq) model (using the PLHNC closure and PW free energy expression) plotted against IUPAC Congener number[350] for an external test set of polychlorobiphenyls (PCBs).

The whole set of PCBs may be separated with respect to the number of chlorine atoms (shown on the top and using dashed vertical lines). It is evident that the error of data predicted with 1D RISM/SDC(QMq) model is constant for all PCBs, whereas the error of solvation continuum models (SM$_6$ and COSMO-SAC) increases with the increase of the number of chlorine atoms in biphenyl. Reprinted with permission from Ref.[29] Copyright 2011 American Chemical Society.

the HFEs calculated by the 3D RISM/UC model are in good agreement with the corresponding experimental data ($R = 0.94$, RMSE=1.06 kcal/mol, ME=-0.72 kcal/mol).

An external test set of 220 chlorinated aromatic hydrocarbons from the list of persistent organic pollutants (polychlorinated benzenes and polychlorinated biphenyls)[29] were used to benchmark the predictive accuracy of the 1D RISM/SDC(QMq) model for HFE calculation. The model gives good results for polychlorinated benzenes (ME and STDE are 0.02 kcal/mol and 0.36 kcal/mol, correspondingly). At the same time, it gives somewhat worse results for polychlorobiphenyls (PCBs) with a systematic bias -0.72 kcal/mol but small standard deviation equals 0.55 kcal/mol. Au-

thors[29] stressed that the error remains the same for the whole set of PCBs. They noted that, in the case of lighter PCB congeners, the chlorine atoms are well-separated from each other, and the total effect of chlorine atoms interactions with the solvent molecules can be represented as a sum of single chlorine atom contributions. Increasing the number of chlorine atoms in biphenyl leads to the interference of the chlorine atoms' interactions with the solvent molecules and, as a result, to a nonlinearity of the solvent response in the process of hydration. The 1D RISM approach considers these effects in a proper way, even in the case of higher chlorinated compounds. In turn, the continuum solvent models (SM_6 and COSMO-SAC)[350] are not sensitive to the nonlinear solvent response (see Figure 19). These results illustrate that the 1D RISM/SDC(QMq) approach is a promising method that can be used to describe hydration/solvation processes for a wide range of chemical solutes.

Accurate computational methods to predict the solubility of organic molecules in water are highly sought after in many fields of the biomolecular sciences.[351,352] For example, predictions of solubility are used in the pharmaceutical and agrochemical industries to assess the bioavailability of *de novo* designed drugs and the environmental fate of potential pollutants, respectively.[353] While a large number of QSPR solubility models have been built by many different research groups, development of a method based on chemical theory and simulation has proved challenging. To address this problem, it has recently been shown that the intrinsic aqueous solubility of crystalline druglike molecules can be estimated with reasonable accuracy from sublimation free energies calculated using crystal lattice simulations and hydration free energies calculated using the 3D RISM/UC method.[12] The solubilities of 25 crystalline druglike molecules taken from different chemical classes were predicted with a R = 0.85 and RMSE=1.45 $\log_{10}$ S units, which was more accurate than the results obtained using implicit continuum solvent models. Unlike QSPR approaches, the model is not directly parametrized against experimental solubility data, and it provides a full computational characterization of the thermodynamics of transfer of the molecule from crystal phase to gas phase to dilute aqueous solution.

The total SFE of a solute as calculated by 3D RISM can be decomposed into a linear sum of

substructure-based contributions using a Spatial Decomposition Analysis (SDA).[311] SDA has been demonstrated to be useful for interpreting the relative solvation of residues in a protein,[311] but may also have many other uses,[7,194,195] including an alternative approach to the semi-empirical models discussed previously.

7 Note on alternative distribution-function approaches

Before closing the review, it will be beneficial to introduce another set of distribution-function approaches to the solvation free energy. Firstly, we would like to attract the readers to recent developments in classical DFT for molecular fluids, a 'sister' method to RISM and MOZ.[164,325–327,354?–366]

The MDFT method developed by Borgis and co-workers is based on classical DFT and MOZ-like six-dimensional representation of positional and orientational solvent density function.[163,164,325–327,365,367,368] The MDFT theory is formulated for an ensemble of molecules where each solvent molecule is considered as a rigid body and characterized by its position and orientation. The solute is also rigid and fixed, it produces an external potential exerted on solvent molecules. As an input MDFT requires (i) chemical structure of the solute and (if applicable) co-solute molecules, (ii) solute-solvent interaction potentials (force-field) and the partial charge distribution of the solute, (iii) bulk properties of the solvent (the structure factor and the k-dependent longitudinal and transverse dielectric constants) as well as 6D direct correlation functions of the uniform solvent that are obtained from MD or MC simulations.[327] The theory can be combined with the so-called homogeneous reference fluid (HRF) approximation.[318,326] In this case the unknown excess free energy of the solute-solvent system can be determined from the angular-dependent direct correlation function of the bulk solvent that can be found by an independent molecular simulation of a box of bulk solvent. By benchmarking the MDFT results against MD simulations, Borgis and co-workers found that such approach is able to provide accurate results for different solute-solvent systems.[163,164,325–327] The first attempt to perform massive SFE calculations (for $\sim$ 500 molecules) by MDFT combined with the PMV-based thermodynamic ensemble correction (see Section 5.5) demonstrates the high potential
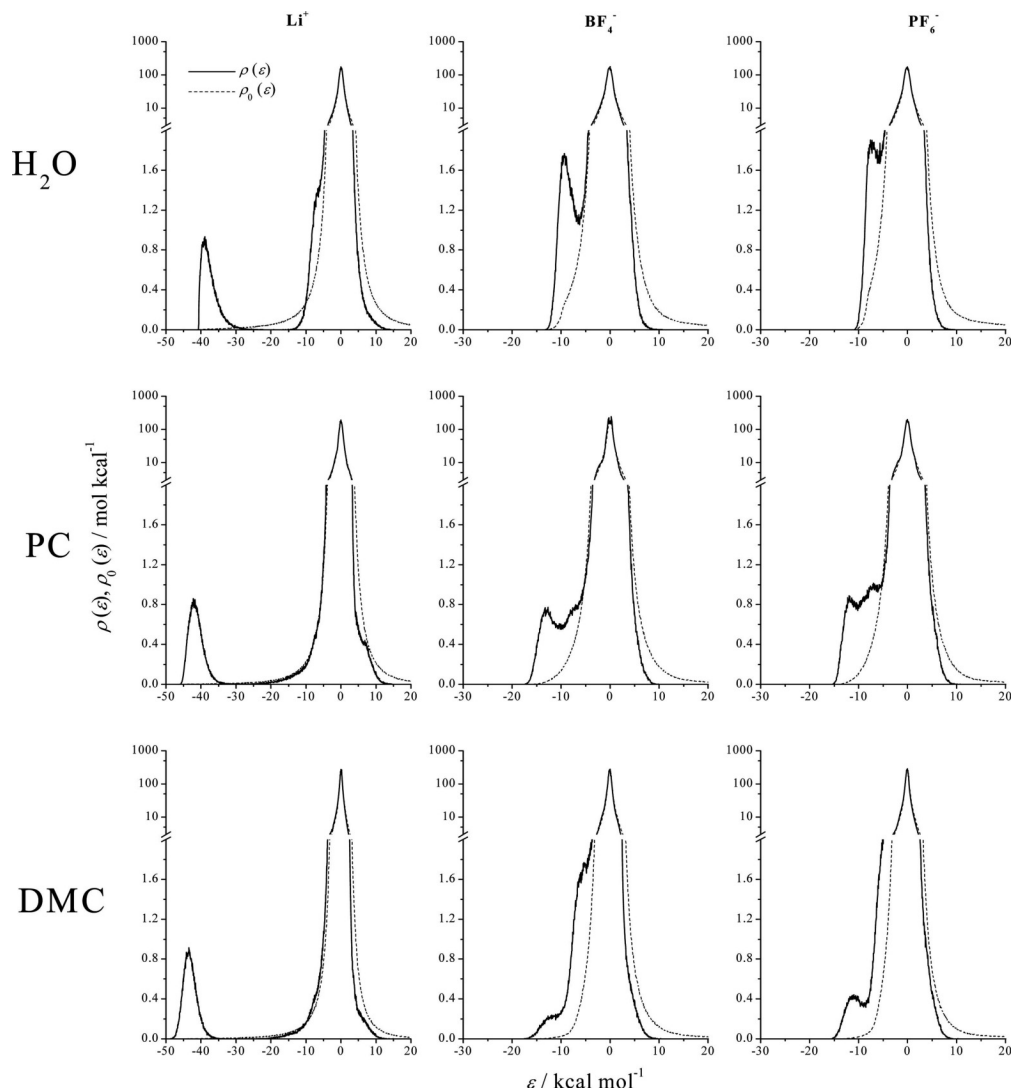
Figure 20: Distribution functions $\rho(\varepsilon)$ and $\rho_0(\varepsilon)$ of the pair interaction energy ε of single ions (Li$^+$, BF$_4^-$ and PF$_6^-$) dissolved in polar solvents (water, PC and DMC). Note a break of the ordinate at 2 mol/kcal. The ordinate is linearly graduated below this value, and is logarithmically graduated above it. $\rho(\varepsilon)$ is the distribution of the pair interaction energy ε in the solution system of interest, while $\rho_0(\varepsilon)$ is the density of states of the solute–solvent potential function multiplied by the bulk solvent density (see Refs[376,384] for details). Reprinted with permission from Ref.[384] Copyright 2012 American Chemical Society.

of the method to be efficiently applied to the problem of accurate prediction of thermodynamic parameters of molecule solvation.[318]

Recent work of Borgis and co-workers presented an extended version of MDFT, MDFT-HNC+3B, that goes beyond the commonly used quadratic expansion of the Gibbs free energy (or equiva-

lently of the excess functional) around the homogeneous reference fluid, and, consequently, be-
yond the HNC approximation.[365] The MDFT-HNC+3B theory imposes a second local minimum
to the Gibbs free energy of the system at low fluid density. The bridge functional that was pro-
posed in this work has several advantages because it "(i) enforces the tetrahedral order of water,
(ii) recovers the close coexistence between gas and liquid states and their surface tension, and (iii)
is consistent with the experimental pressure of the fluid."[365] The MDFT-HNC+3B approach was
benchmarked on a set of reference systems (linear alkanes in water). It was shown that with use of
only one empirical parameter the MDFT-HNC+3B method reproduces results of explicit simula-
tions with a small systematic error of the order of $k_B T$ that is presumably the most accurate result
for MDFT-like theories up to date.[365]

Wu and co-workers have recently developed another version of classical DFT for molecule sol-
vation with a focus on aqueous systems.[355,359?,360] In the so-called Site Density Functional Theory
(SDFT) model the excess Helmholtz energy functional for the solute-solvent system is constructed
as a quadratic expansion with respect to the local density deviation from that of a uniform system;
to resolve the problem of evaluating higher-order terms there has been introduced a universal func-
tional that approximates all higher-order terms by an analytic solution for a reference hard-sphere
system.[359,360] Similar to the version of MDFT developed by Borgis and co-workers,[325–327] as an
input, the SDFT requires independent evaluation of atomistic pair direct correlation functions of
the uniform solvent system that can be calculated from MD simulations. The theory has been suc-
cessfully applied to solvation of amino-acids in water[359] and high-throughout prediction of HFEs
for small molecules.[360] In the latter work it has been shown that the SDFT with the AMBER force-
field for the solutes the TIP3P model for water can predict HFE of 500 neutral molecules with
average unsigned errors of around 1 kcal/mol in comparison with the corresponding experimen-
tal and MD simulation data.[360] Recently, this theory was used by Fu, Liu and Wu for thorough
benchmarking of different DFT-based multi scale methods for predicting HFE of small solutes.[366]

Chong and Ham recently developed a formally exact integral equation for the three-dimensional
hydration structure around molecular solutes of arbitrary complexity, aqueous interaction site

(AXIS) theory.[369] The derivation of the AXIS equation is based on classical DFT formalism for non-uniform polyatomic fluid developed by Chandler, McCoy and Singer.[370,371] The AXIS theory fully takes into account the intramolecular structural correlations of solvent water and provides results on solute-solvent structure for a water solute at ambient and supercritical conditions that are more consistent with molecular simulations than the standard 3D RISM/HNC method.[369] However, the theory in its current form is only applicable to water and there still remain quantitative differences from the simulation results.[369] Nevertheless, we think that the AXIS approach is promising because it provides a formally exact expression for both intramolecular and intermolecular bridge functions;[369] therefore, it can be used as theoretical basis for future developments of accurate moleculare theories.

In RISM and classical DFT methods discussed above, the argument of the distribution functions is distance or spatial coordinate. Instead, Matubayasi and Nakahara put forward an approach using the solute-solvent pair interaction energy.[126] This approach is called the method of energy representation, and the functional for the solvation free energy is constructed from the average distribution function $\rho(\varepsilon)$ of the solute-solvent pair energy ε in the solution of interest, the density of states of solute-solvent pair potential, and the solvent-solvent pair correlation represented over the energy coordinate.[372,373]

Figure 20 illustrates typical behaviour of solute–solvent pair energy distribution functions. The figure shows these functions for several single ions (Li^+, BF_4^- and PF_6^-) dissolved in different polar solvents (water, propylene carbonate (PC) and dimethyl carbonate (DMC)).[384] The negative and positive regions of ε correspond to the attractive and repulsive interactions between solute and solvent, respectively. All these functions have a strong peak around zero due to a large number of solute-solvent pairs with large distances. Other peaks in the regions of negative ε reveal the attractive solute-solvent pair energies. For instance, the peak at 38 kcal/mol for $\rho(\varepsilon)$ of Li^+ in water originates from the strong attraction of water molecules in the first solvation shell to the ion; correspondingly, the shoulder at the negative envelope near the intense peak originates from the ion-water interactions for water molecules in the outer solvation shells of the ion.[384]

As is so for the RISM approaches, the argument of distribution functions is a projected coordinate of solute-solvent configuration, which is of high dimension. In 1D and 3D RISM, the projected coordinates are the distance and the spatial coordinate, respectively, and the dimensionality reduction to 1 or 3 dimensions is carried out. In the method of energy representation, the argument of the distribution functions is the solute-solvent pair energy. This is another form of projection, and the projected coordinate is always of 1 dimension. The basic idea behind the projection onto energy lies in the statistical-mechanics principle that only the energy is extracted from any configuration (structure) in the ensemble to determine the system free energy. The energy-representation method sticks to this principle on the solute-solvent pair level. Matubayasi and Nakahara provided a DFT-type formulation toward the solvation free energy in terms of distribution functions of solute-solvent pair energy, without many sophisticated approximations.[126,372,373] Still, the above principle in statistical mechanics cannot be violated on the solute-solvent pair level since any solute-solvent pair configurations with equal interaction energies contribute to the solvation free energy automatically in equal weights.

In the energy-representation method, the problem of auxiliary site is absent and the free energy functional is exact to second order in solvent density. The molecular flexibility and the system inhomogeneity are handled without any modification of the method since the projection of solute-solvent configuration is done in terms of energy; the flexibility and inhomogeneity are spatial information, which are to be projected out in the formulation. Through combination with molecular simulation, the error due to the approximation in the method is observed to be not larger than the error due to the use of force field,[317] though the computational speed is inferior to those of (1D) RISM-based schemes. The energy-representation method provides a universal computational and theoretical platform for calculation of thermodynamic parameters of solvation and has been already applied to a range of different systems: supercritical fluids, air-water interfaces, micelles, lipid membranes, proteins, and QM/MM systems.[129,374–383,385] The approximation used in the current version of energy-representation method is simply PY-type and HNC-type ones over the energy coordinate, however, it is considered a new class of bridge functional since the coordi-

nate (abscissa) of distribution function has been switched from distance to energy. The connection between spatial coordinate-based and energy-based approaches can be an interesting topic of investigation.

8   Conclusion

As the critical analysis of the literature above shows, there have been many significant developments in molecular theories since the mid 2000s, which have resulted in increased accuracy of these methods and increased ease-of-use for non-specialists. This has led to a rapid growth of different application areas of the theories. The field has significantly matured in terms of its applications and it is safe to say that now practically all areas of molecular and biomolecular sciences are covered by integral equitation theories and their 'sister' methods like classical DFT and energy-representation theories. These theories provide a theoretical framework for calculating the structural and thermodynamic properties of liquid-phase solutions. They are applicable to both pure and mixed solvent systems and can be applied for a vast range of solute-solvent systems.

Particularly impressive achievements have been made in applications of these theories to various aqueous solutions - from single ion solvation to nano-porous systems and protein-drug binding. Non-aqueous systems attracted somewhat less attention; however, the situation is currently changing as evidenced by the ever increasing number of applications of molecular theories in ionic liquids, supercritical fluids and organic solvents.

Until recently, molecular theories have been mostly considered as *qualitative* tools; however the recent developments in the field discussed in the review have made it possible to use these methods for making accurate *quantitative* predictions of structural and thermodynamic properties of various solute-solvent systems. We note though that the majority of these high-accuracy predictions were done for systems at ambient conditions and there is still a number of challenges to overcome (accurate predictions of properties at non-ambient conditions, development of high-accuracy methods for arbitrary selected non-aqueous solvents, highly concentrated electrolytes etc). Nevertheless, to

our opinion, the prospects of further improvements of the prediction accuracy of molecular theories seem to be optimistic. In particular we expect a significant progress in the field during the next decade related with (i) development of effective codes for high-dimensional theories (like MOZ); (ii) development of general accurate bridge functionals; (iii) incorporation of solvent dynamics in the theories; (iv) linking of molecular theories with other molecular modeling methods (e.g. QM and MM) and cheminformatics tools; (v) extensive benchmarking of the theories by new sets of experimental data.

We expect an increased interest of industrial applications of RISM and other molecular theories in the nearest future that would include fragment-based drug design (pharma-related areas), calculating protein-ligand binding free energies (pharma-related areas, cosmetics and food industry), optimising properties of energy storage devices such as super-capacitors (energy-related areas) and modelling of fluid properties in porous systems (chemical engineering, oil&gas-related areas).

It is also anticipated that the popularity of molecular theories will rapidly increase in the near future, driven by the recent implementation of of 1D and 3D RISM methods in the widely used Amber, MOE and ADF molecular modelling packages.

9   Acknowledgment

10 Acronyms

1D one dimensional

3D three dimensional

3D-SFED 3D Solvation Free Energy Density

6D six dimensional

A^- acidic anion

AcrB Acriflavine resistance protein B

ADF Amsterdam Density Functional

Amber Assisted Model Building with Energy Refinement

AQP1 Aquaporin 1

ASFE advanced solvation force extrapolation

AXIS aqueous interaction site (theory)

B3LYP Becke, three-parameter, Lee-Yang-Parr

COSMO-SAC conductor-like screening model – segment activity coefficient

CPU Central Processing Units

DDB Dortmund Databank; http://www.ddbst.com/ddb.html

DFT density functional theory

DHH Duh-Haymet-Henderson

DIIS direct inversion of the iterative subspace

DMC dimethyl carbonate

DMSO dimethyl sulfoxide

DNA Deoxyribonucleic acid

DPMV dimensionless partial molar volume

DRISM dielectrically consistent reference interaction sites model

EC-RISM embedded cluster reference interaction site model

EPICS equilibrium partition in closed system

FEP free energy perturbation

GC group contribution

GF Gaussian fluctuations

GlpF glycerol facilitator

GMRES Generalized minimum residual method

GPU Graphics Processing Unit

GSM gas stripping method

H^+ proton

HA monoprotic acid

HF Hartree-Fock

HFE hydration free energy

HNC hypernetted chain

HNCB hypernetted chain closure with repulsive bridge correction functions

IET integral equation theory

ISc Initial State Correction

KH Kovalenko–Hirata

L ligand

LJ Lennard–Jones

MC Monte Carlo

MCSCF multiconfigurational self-consistent field

MD molecular dynamics

MDIIS modified direct inversion of the iterative subspace

ME mean of error

Met-enkephalin [Methionine]enkephalin

MHNC modified hypernetted chain

MM molecular mechanics

MOE Molecular Operating Environment

MOZ molecular Ornstein–Zernike

MS Martynov-Sarkisov

MSA mean spherical approximation

MSV Martynov-Sarkisov-Vompe

MTS-MD multiscale method of multiple time step molecular dynamics

NAB nucleic acid builder

NEW non-equilibrium work

NMR Nuclear Magnetic Resonance

NR Newton-Raphson

OIN optimized isokinetic Nosé-Hoover chain thermostat

OPLS optimized potential for liquid simulations

OZ Ornstein–Zernike

P partition coefficient (Attention, in Eq.6 P stays for protein)

PC propylene carbonate

PCB polychlorobiphenyl

PCM polarizable continuum model

PL protein-ligand complex

PLHNC partially linearized hypernetted chain

PMF potential of mean force

PMV partial molar volume

POP persistent organic pollutant

PRV phase ratio variation

PW partial wave

PWC partial wave correction

PY Percus-Yevick

QM quantum mechanical or quantum mechanics

QSPR quantitative structure - property relationships

RDF radial distribution function

RHF restricted Hartree-Fock

RHNC reference hypernetted chain

RISM reference interaction sites model

RMSE root mean square of error

RNT rosette nanotubes

SAMPL Statistical Assessment of Modeling of Proteins and Ligands

SASA solvent accessible surface area

SCF self-consistent field

SDC structural descriptors correction

SEDD spatial electron density distribution

SFE solvation free energy

SM6 solvation model No. 6

SMSA soft core mean spherical approximation

SPC/E simple point charge/extended

SSOZ site-site Ornstein–Zernike

STDE standard deviation of error

TEC Thermodynamic Ensemble Correction

TI thermodynamic integration

TIP3P transferable intermolecular potential, three-point

TIP4Pew transferable intermolecular potential, four-point, for use with Ewald summation methods

UC universal correction

WWC wetted-wall column

XRISM extended reference interaction sites model

11 Appendix

Appendix 1

Khalatur's scheme of 1D RISM–MC calculations:[289]

    Step 0 (Preparation step)

a) Prepare an initial chain configuration, $\mathbf{r}_{A0} \equiv \{\mathbf{r}_1, ..., \mathbf{r}_N\}$.

b) Calculate the solvent susceptibility function $\chi_{\alpha\beta}(r)$ with previously evaluated functions $\omega_{\alpha\beta}^{solv}(r)$ and $h_{\alpha\beta}^{solv}(r)$ (see eq 35), determine its Fourier transform in q-space $\widehat{\chi}_{\alpha\beta}(q)$.

c) Set the potential $\Delta\Psi(\mathbf{r})$ to zero.

d) Set the iteration counter τ to zero.

Step 1

a) Generate consistently a set of k chain configurations by the MC conformation annealing; accept or reject the 'move' $\mathbf{r}_{A0} \to \mathbf{r}_{Ai}$ ($i = 1 \ldots k$) according to the Metropolis criterion:

$$P_{accept} \propto \exp[-\beta(\Psi_i(\mathbf{r}) - \Psi_{i+1}(\mathbf{r}))],$$

where $\Psi(\mathbf{r})$ is determined by eq 56.

b) Calculate intramolecular correlation functions $\omega_{ss'}(r)$ (see eq 32) for the latest configuration $\mathbf{r}_{Ak}$.

c) Repeat operations (1.a) and (1.b) m times and calculate an intermediate averaged intramolecular correlation functions:

$$(\omega_{ss'}(r))_{av} = \frac{1}{m}\sum_{i=1}^{m}(\omega_{ss'}(r))_i,$$

calculate their Fourier transforms $(\widehat{\omega}_{ss'}(q))_{av}$.

Step 2

a) Use the functions $(\widehat{\omega}_{ss'}(q))_{av}$ in the Fourier q-space to solve RISM equation (see eq 34) together with a closure relation (see eq 36); obtain correlation functions $\widehat{c}_{s\alpha}(q)$ and $\widehat{h}_{s\alpha}(q)$.

b) Evaluate the solvent-mediated potential $\Delta\widehat{\Psi}_{ss'}(q)$ by substitution of $\widehat{c}_{s\alpha}(q)$ and $\widehat{\chi}_{\alpha\beta}(q)$ into eq 58.

c) Calculate $\Delta\widehat{\Psi}(\mathbf{q})$ using eq 57, transform it to $\Delta\Psi(\mathbf{r})$.

Step 3

a) Increase the iteration counter τ by unity: $\tau=\tau+1$.

b) Calculated the current averaged solvent-mediated potential:

$$(\Delta\Psi(\mathbf{r}))_{av} = \frac{1}{\tau}\sum_{i=1}^{\tau}\Delta\Psi_i(\mathbf{r})$$

c) Evaluate the achieved accuracy. Return to **Step 1** if the required accuracy is not obtained (use the obtained potential $(\Delta\Psi(r))_{av}$ instead of the previous one in operation (1.a)).

Appendix 2

Kast's 1D RISM–MC technique to study the complexation thermodynamics:[236]

1. The set of equations: 1D RISM equation (eq 34), closure relation (eq 36), and solvation free energy functional (see the Section End-point SFE functionals in 1D RISM), is solved to yield solvation free energies of the host molecule ($\Delta\mu_X$) and the guest molecule ($\Delta\mu_Y$).

2. A supramolecular structure is constructed by placing the host molecule (X) and the guest molecule (Y) at a large distance. The same set of equations is solved to obtain a set of distribution functions $\mathbf{h}_0$ and $\mathbf{c}_0$. The offset of the PMF is calculated as follows:

$$W(r,\theta,\varphi) = \Delta U_{XY}(r,\theta,\varphi)+$$
$$\Delta\mu_{XY}(r,\theta,\varphi) - \Delta\mu_X - \Delta\mu_Y,$$

where $\Delta U_{XY}(r,\theta,\varphi)$ is the vacuum host–guest interaction energy, θ and φ are the angular coordinates.

3. A set of host–guest configurations is generated randomly by variation of the angular coordinates. The PMF difference $W(r',\theta,\varphi) - W(r,\theta,\varphi)$ for a small step of the guest toward the

host is calculated.

4. Configurations are accepted with the probability P meeting the Metropolis criterion:

$$P_{accept} \propto \exp[-\beta(W(r', \theta, \varphi) - W(r, \theta, \varphi)].$$

5. The 1D PMF difference $\omega(r') - \omega(r)$ for a step of the guest toward the host is sampled as follows:

$$\omega(r') - \omega(r) = -k_B T \ln \langle A \rangle_r,$$

where $A = \exp[-\beta(W(r', \theta, \varphi) - W(r, \theta, \varphi)]$.

6. Steps 3–5 are repeated until the reaction coordinate approaches zero.

Appendix 3

Benchmarks of the RISM calculations: As we mentioned above, 1D and 3D RISM were implemented in two widely used software packages, ADF[10] and Amber.[1] For newcomers we may recommend the following benchmarks of the RISM calculations. Frolov et al[386] used the NAB simulation package[1] in the AmberTools 1.4 set of routines to perform the 3D RISM calculations.[387] The following 3D RISM parameters were investigated: *Buffer* - the smallest distance between solute atoms and a 3D box side; *spacing* - the distance between grid points in 3D-grid (see Figure 8). The tolerance (the L_2 norm of the difference between two subsequent solutions of 3D RISM iterations) was set to 10^{-10} and the number of vectors used by MDIIS solver was 5 and MDIIS step size - 0.7 following the works of the developers of the 3D RISM in the AmberTools.[1,387] Paracetamol was selected as the test solute to perform these benchmarks. Satisfactory reproduceability (error in a range of 0.02 kcal/mol) could be obtained using the following parameters: buffer = 30 Å, spacing = 0.3 Å. The solvent susceptibility functions for 3D RISM calculations were obtained by the 1D RISM method present in the AmberTools 1.4. The dielectrically consistent 1D RISM (DRISM) was employed[222] with the PLHNC closure.[205] The grid size for 1D-functions was 0.025

Å, which gave a total of 16384 grid points. We employed the MDIIS iterative scheme,[242] where we used 20 MDIIS vectors, MDIIS step size - 0.3, and residual tolerance - 10^{-12}. The solvent was considered to be pure water with the number density 0.0333 Å^{-3}, a dielectric constant of 78.497, at a temperature of 300K.

12   References

(1) Luchko, T.; Gusarov, S.; Roe, D. R.; Simmerling, C.; Case, D. A.; Tuszynski, J.; Kovalenko, A. Three-Dimensional Molecular Theory of Solvation Coupled with Molecular Dynamics in Amber. *J. Chem. Theory Comput.* **2010**, *6*, 607–624.

(2) Frolov, A. I.; Ratkova, E. L.; Palmer, D. S.; Fedorov, M. V. Hydration Thermodynamics using the Reference Interaction Site Model: Speed or Accuracy? *J. Phys. Chem. B* **2011**, *115*, 6011–6022.

(3) Kovalenko, A.; Kobryn, A. E.; Gusarov, S.; Lyubimova, O.; Liu, X.; Blinov, N.; Yoshida, M. Molecular theory of solvation for supramolecules and soft matter structures: application to ligand binding, ion channels, and oligomeric polyelectrolyte gelators. *Soft Matter* **2012**, *8*, 1508–1520.

(4) Sergiievskyi, V. P.; Fedorov, M. V. 3DRISM Multigrid Algorithm for Fast Solvation Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 2062–2070.

(5) Maruyama, Y.; Hirata, F. Modified Anderson Method for Accelerating 3D-RISM Calculations Using Graphics Processing Unit. *J. Chem. Theory Comput.* **2012**, *8*, 3015–3021.

(6) Kovalenko, A. Molecular Theory of Solvation and Electrical Double Layer in Nanoporous Carbon Electrodes. In *Electrochemical Capacitors: Fundamentals to Applications*;

Long, J. W., Brousse, T., Simon, P., Belanger, D., Sugimoto, W., Brodd, R., Eds.; 2012; Vol. 41; pp 133–149.

(7) Kovalenko, A. Multiscale modeling of solvation in chemical and biological nanosystems and in nanoporous materials. *Pure Appl. Chem.* **2013**, *85*, 159–199.

(8) Maruyama, Y.; Yoshida, N.; Tadano, H.; Takahashi, D.; Sato, M.; Hirata, F. Massively parallel implementation of 3D-RISM calculation with volumetric 3D-FFT. *J. Comput. Chem.* **2014**, *35*, 1347–1355.

(9) Misin, M.; Fedorov, M. V.; Palmer, D. S. Accurate hydration free energies at a wide range of temperatures from 3D-RISM. *J. Chem. Phys.* **2015**, *142*, 091105(1-6).

(10) Gusarov, S.; Ziegler, T.; Kovalenko, A. Self-consistent combination of the three-dimensional RISM theory of molecular solvation with analytical gradients and the Amsterdam density functional package. *J. Phys. Chem. A* **2006**, *110*, 6083–6090.

(11) Molecular Operating Environment (MOE), 2013.08; Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite 910, Montreal, QC, Canada, H3A 2R7. 2013; http://www.chemcomp.com/. Accessed 23rd September 2014.

(12) Palmer, D. S.; McDonagh, J. L.; Mitchell, J. B. O.; van Mourik, T.; Fedorov, M. V. First-Principles Calculation of the Intrinsic Aqueous Solubility of Crystalline Druglike Molecules. *J. Chem. Theory Comput.* **2012**, *8*, 3322–3337.

(13) Imai, T.; Oda, K.; Kovalenko, A.; Hirata, F.; Kidera, A. Ligand Mapping on Protein Surfaces by the 3D-RISM Theory: Toward Computational Fragment-Based Drug Design. *J. Am. Chem. Soc.* **2009**, *131*, 12430–12440.

(14) Imai, T.; Miyashita, N.; Sugita, Y.; Kovalenko, A.; Hirata, F.; Kideras, A. Functionality Mapping on Internal Surfaces of Multidrug Transporter AcrB Based on Molecular Theory

of Solvation: Implications for Drug Efflux Pathway. *J. Phys. Chem. B* **2011**, *115*, 8288–8295.

(15) Phongphanphanee, S.; Yoshida, N.; Hirata, F. Molecular Recognition Explored by a Statistical-Mechanics Theory of Liquids. *Curr. Pharm. Des.* **2011**, *17*, 1740–1757.

(16) Imai, T.; Kinoshita, M.; Hirata, F. Salt effect on stability and solvation structure of peptide: An integral equation study. *Bull. Chem. Soc. Jpn.* **2000**, *73*, 1113–1122.

(17) Imai, T.; Hiraoka, R.; Kovalenko, A.; Hirata, F. Water molecules in a protein cavity detected by a statistical-mechanical theory. *J. Am. Chem. Soc.* **2005**, *127*, 15334–15335.

(18) Yoshida, N.; Phongphanphanee, S.; Maruyama, Y.; Imai, T.; Hirata, F. Selective ion-binding by protein probed with the 3D-RISM theory. *J. Am. Chem. Soc.* **2006**, *128*, 12042–12043.

(19) Howard, J. J.; Lynch, G. C.; Pettitt, B. M. Ion and Solvent Density Distributions around Canonical B-DNA from Integral Equations. *J. Phys. Chem. B* **2011**, *115*, 547–556.

(20) Phongphanphanee, S.; Yoshida, N.; Oiki, S.; Hirata, F. Distinct configurations of cations and water in the selectivity filter of the KcsA potassium channel probed by 3D-RISM theory. *J. Mol. Liq.* **2014**, *200, Part A*, 52–58.

(21) Kast, S. M.; Heil, J.; Gussregen, S.; Schmidt, K. F. Prediction of tautomer ratios by embedded-cluster integral equation theory. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 343–353.

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

(23) Miyata, T.; Hirata, F. Combination of molecular dynamics method and 3D-RISM theory for conformational sampling of large flexible molecules in solution. *J. Comput. Chem.* **2008**, *29*, 871–882.

(24) Chong, S. H.; Ham, S. Configurational entropy of protein: A combined approach based on molecular simulation and integral-equation theory of liquids. *Chem. Phys. Lett.* **2011**, *504*, 225–229.

(25) Chong, S. H.; Lee, C.; Kang, G.; Park, M.; Ham, S. Structural and Thermodynamic Investigations on the Aggregation and Folding of Acylphosphatase by Molecular Dynamics Simulations and Solvation Free Energy Analysis. *J. Am. Chem. Soc.* **2011**, *133*, 7075–7083.

(26) Omelyan, I.; Kovalenko, A. Generalised canonical-isokinetic ensemble: speeding up multi-scale molecular dynamics and coupling with 3D molecular theory of solvation. *Mol. Simul.* **2013**, *39*, 25–48.

(27) Omelyan, I.; Kovalenko, A. Multiple time step molecular dynamics in the optimized isokinetic ensemble steered with the molecular theory of solvation: Accelerating with advanced extrapolation of effective solvation forces. *J. Chem. Phys.* **2013**, *139*, 244106(1-23).

(28) Ben-Naim, A. *Molecular Theory of Solutions*; Oxford University Press, USA, 2006.

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

(30) Collins English Dictionary - Complete & Unabridged 10th Edition. 2011; http://dictionary.reference.com/browse/ partitioncoefficient. Accessed 19th December 2013.

(31) One should note that the Henry's law constants can be defined in several ways which fall in two groups: constants representing the volatility (i.e. gas-phase composition divided by aqueous-phase composition) and inverse constants representing the solubility (i.e. aqueous-phase composition divided by gas-phase composition). See *www.henrys-law.org* for more information. In the current manuscript eq 1 corresponds to the definition from the first group,

where K_H is expressed as the *dimensionless* ratio between the equilibrium molecular concentrations. We need to underline that the Henry's law constants can be expressed using different units ($m^3*Pa*mol^{-1}$, atm, dimensionless, etc.), which are linked one to another by conversion factors (see *www.henrys-law.org*).

(32) Converting Henry's Law Constants. http://www.mpch-mainz.mpg.de/ sander/res/henry-conv.html. Accessed 19th December 2013.

(33) Mackay, D.; Shiu, W. Y.; Ma, K. C. *Illustrated Handbook of Physical-Chemical Properties and Environmental Fate for Organic Chemicals, Volume 2, Polynuclear Aromatic Hydrocarbons, Polychlorinated Dioxins, and Dibenzofurans*; Lewis Publishers, 1992.

(34) Valsaraj, K. T.; Thibodeaux, L. J. On the Physicochemical Aspects of the Global Fate and Long-Range Atmospheric Transport of Persistent Organic Pollutants. *J. Phys. Chem. Lett.* **2010**, *1*, 1694–1700.

(35) Gong, S. L.; Huang, P.; Zhao, T. L.; Sahsuvar, L.; Barrie, L. A.; Kaminski, J. W.; Li, Y. F.; Niu, T. GEM/POPs: a global 3-D dynamic model for semi-volatile persistent organic pollutants - Part 1: Model description and evaluations of air concentrations. *Atmos. Chem. Phys.* **2007**, *7*, 4001–4013.

(36) Strand, A.; Hov, O. A model strategy for the simulation of chlorinated hydrocarbon distributions in the global environment. *Water Air Soil Pollut.* **1996**, *86*, 283–316.

(37) Liss, P. S.; Slater, P. G. Flux of Gases Across Air-sea Interface. *Nature* **1974**, *247*, 181–184.

(38) Vallack, H. W. et al. Controlling persistent organic pollutants - what next? *Environ. Toxicol. Pharmacol.* **1998**, *6*, 143–175.

(39) Jones, K. C.; de Voogt, P. Persistent organic pollutants (POPs): state of the science. *Environ. Pollut.* **1999**, *100*, 209–221.

(40) Wine, P. H. Atmospheric and Environmental Physical Chemistry: Pollutants without Borders. *J. Phys. Chem. Lett.* **2010**, *1*, 1749–1751.

(41) Aarhus Protocol on Persistent Organic Pollutants (POPs). http://www.unece.org/env/ lrtap/pops_h1.htm. Accessed 19th December 2013.

(42) Fujita, T.; Hansch, C.; Iwasa, J. A New Substituent Constant Pi Derived From Partition Coefficients. *J. Am. Chem. Soc.* **1964**, *86*, 5175–5180.

(43) Takács-Novák, K.; Avdeef, A. Interlaboratory study of log P determination by shake-flask and potentiometric methods. *J. Pharm. Biomed. Anal.* **1996**, *14*, 1405–1413.

(44) Avdeef, A.; Box, K. J.; Comer, J. E. A.; Hibbert, C.; Tam, K. Y. pH-metric logP 10. Determination of liposomal membrane-water partition coefficients of ionizable drugs. *Pharm. Res.* **1998**, *15*, 209–215.

(45) Smith, G. D.; Borodin, O.; Magda, J. J.; Boyd, R. H.; Wang, Y. S.; Bara, J. E.; Miller, S.; Gin, D. L.; Noble, R. D. A comparison of fluoroalkyl-derivatized imidazolium:TFSI and alkyl-derivatized imidazolium:TFSI ionic liquids: a molecular dynamics simulation study. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7064–7076.

(46) Rösgen, J.; Hinz, H.-J. Theory and practice of DSC measurements on proteins, In *From Macromolecules to Man*; Kemp, R., Ed.; Handbook of Thermal Analysis and Calorimetry; Elsevier Science B.V., 1999; Vol. 4; pp 63–108.

(47) Podjarny, A.; Dejaegere, A. P.; Kieffer, B. *Biophysical Approaches Determining Ligand Binding to Biomolecular Targets: Detection*; 2011.

(48) Perozzo, R.; Folkers, G.; Scapozza, L. Thermodynamics of protein-ligand interactions: history, presence, and future aspects, *J. Recept. Signal Transduct. Res.* **2004**, *24*, 1–52.

(49) Zevatskii, Y. E.; Samoilov, D. V.; Mchedlov-Petrosyan, N. O. Conthemporary methods for

the experimental determination of dissociation constants of organic acids in solutions. *Russ. J. Gen. Chem.* **2009**, *79*, 1859–1889.

(50) Duc, M.; Gaboriaud, F.; Thomas, F. Sensitivity of the acid-base properties of clays to the methods of preparation and measurement - 1. Literature review. *J. Colloid Interface Sci.* **2005**, *289*, 139–147.

(51) Alongi, K. S.; Shields, G. C. Theoretical Calculations of Acid Dissociation Constants: A Review Article. In *Annual Reports in Computational Chemistry*; Wheeler, R. A., Ed.; Elsevier, 2010; Vol. 6; pp 113 – 138.

(52) Palmer, D. S.; Llinas, A.; Morao, I.; Day, G. M.; Goodman, J. M.; Glen, R. C.; Mitchell, J. B. O. Predicting intrinsic aqueous solubility by a thermodynamic cycle. *Mol. Pharmaceutics* **2008**, *5*, 266–279.

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

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

(55) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. Predicting aqueous solubilities from aqueous free energies of solvation and experimental or calculated vapor pressures of pure substances. *J. Chem. Phys.* **2003**, *119*, 1661–1670.

(56) Ben-Amotz, D.; Underwood, R. Unraveling Water's Entropic Mysteries: A Unified View of Nonpolar, Polar, and Ionic Hydration. *Acc. Chem. Res.* **2008**, *41*, 957–967.

(57) Krestov, G. A., Novosyolov, N. P., Perelygin, I. S., Kolker, A. M., Safonova, L. P., Ovchinnikova, V. D., Trostin, V. N., Eds. *Ionic solvation*; Ellis Horwood Series in Inorganic Chemistry; Horwood, New York, 1994.

(58) Blumberger, J.; Bernasconi, L.; Tavernelli, I.; Vuilleumier, R.; Sprik, M. Electronic Structure and Solvation of Copper and Silver Ions: A Theoretical Picture of a Model Aqueous Redox Reaction. *J. Am. Chem. Soc.* **2004**, *126*, 3928–3938.

(59) Hamelberg, D.; McCammon, J. A. Standard Free Energy of Releasing a Localized Water Molecule from the Binding Pockets of Proteins: Double-Decoupling Method. *J. Am. Chem. Soc.* **2004**, *126*, 7683–7689.

(60) Baron, R.; Setny, P.; McCammon, J. A. Water in Cavity-Ligand Recognition. *J. Am. Chem. Soc.* **2010**, *132*, 12091–12097.

(61) Hughes, L. D.; Palmer, D. S.; Nigsch, F.; Mitchell, J. B. O. Why are some properties more difficult to predict than others? A study of QSPR models of solubility, melting point, and Log P. *J. Chem. Inf. Model.* **2008**, *48*, 220–232.

(62) Kerns, E. H.; Di, L. *Drug-like properties: concepts, structure design and methods: from ADME to toxicity optimization*; Academic Press, 2008.

(63) Palmer, D. S.; O'Boyle, N. M.; Glen, R. C.; Mitchell, J. B. O. Random forest models to predict aqueous solubility. *J. Chem. Inf. Model.* **2007**, *47*, 150–158.

(64) O'Boyle, N. M.; Palmer, D. S.; Nigsch, F.; Mitchell, J. B. O. Simultaneous feature selection and parameter optimisation using an artificial ant colony: case study of melting point prediction, *Chem. Cent. J.* **2008**, *2*, 21(1-15).

(65) Bamford, H. A.; Baker, J. E. *Review of Methods and Measurements of Selected Hydrophobic Organic Contaminant Aqueous Solubilities, Vapor Pressures, and Air-Water Partition Coefficients*; Gaithersburg, Md.: National Institute of Standards and Technology, 1998; Chapter Physical and Chemical properties.

(66) Lincoff, A. H.; Gossett, J. M. In *Gas Transfer at Water Surface*; Brusasert, W., Jirka, G. H., Eds.; Dordrecht, Holland, 1984.

118

(67) Robbins, G. A.; Wang, S.; Stuart, J. D. Using the Static Headspace Method To Determine Henry Law Constants. *Anal. Chem.* **1993**, *65*, 3113–3118.

(68) Mackay, D.; Shiu, W. Y.; Sutherland, R. P. Determination of Air-water Henrys Law Constants For Hydrophobic Pollutants. *Environ. Sci. Technol.* **1979**, *13*, 333–337.

(69) Fendinger, N. J.; Glotfelty, D. E. A Laboratory Method For the Experimental-determination of Air-water - Henrys Law Constants For Several Pesticides. *Environ. Sci. Technol.* **1988**, *22*, 1289–1293.

(70) Goss, K. U.; Wania, F.; McLachlan, M. S.; Mackay, D.; Schwarzenbach, R. P. Comment on "Reevaluation of air-water exchange fluxes of PCBs in green bay and southern Lake Michigan". *Environ. Sci. Technol.* **2004**, *38*, 1626–1628.

(71) Shunthirasingham, C.; Lei, Y. D.; Wania, F. Evidence of bias in air-water Henry's law constants for semivolatile organic compounds measured by inert gas stripping. *Environ. Sci. Tech.* **2007**, *41*, 3807–3814.

(72) Bamford, H. A.; Baker, J. E.; Poster, D. L. *Review of methods and measurements of selected hydrophobic organic contaminant aqueous solubilities, vapor pressures, and air-water partition coefficients*; NIST special publication 928; U. S. Government Printing Office Washington, 1998; p 95.

(73) Glomme, A.; Marz, J.; Dressman, J. B. Comparison of a miniaturized shake-flask solubility method with automated potentiometric acid/base titrations and calculated solubilities. *J. Pharm. Sci.* **2005**, *94*, 1–16.

(74) Llinas, A.; Burley, J. C.; Prior, T. J.; Glen, R. C.; Goodman, J. M. Concomitant hydrate polymorphism in the precipitation of sparfloxacin from aqueous solution. *Cryst. Growth Des.* **2008**, *8*, 114–118.

(75) Llinas, A.; Glen, R. C.; Goodman, J. M. Solubility challenge: Can you predict solubilities of 32 molecules using a database of 100 reliable measurements? *J. Chem. Inf. Model.* **2008**, *48*, 1289–1303.

(76) Avdeef, A.; Berger, C. M. pH-metric solubility. 3. Dissolution titration template method for solubility determination. *Eur. J. Pharm. Sci.* **2001**, *14*, 281–291.

(77) Stuart, M.; Box, K. Chasing equilibrium: Measuring the intrinsic solubility of weak acids and bases. *Anal. Chem.* **2005**, *77*, 983–990.

(78) Hopfinger, A. J.; Esposito, E. X.; Llinas, A.; Glen, R. C.; Goodman, J. M. Findings of the Challenge to Predict Aqueous Solubility. *J. Chem. Inf. Model.* **2009**, *49*, 1–5.

(79) Rohac, V.; Ruzicka, V.; Ruzicka, K.; Polednicek, M.; Aim, K.; Jose, J.; Zabransky, M. Recommended vapour and sublimation pressures and related thermal data for chlorobenzenes. *Fluid Phase Equilib.* **1999**, *157*, 121–142.

(80) Ruzicka, K.; Fulem, M.; Ruzicka, V. Recommended vapor pressure of solid naphthalene. *J. Chem. Eng. Data* **2005**, *50*, 1956–1970.

(81) Fulem, M.; Ruzicka, K.; Ruzicka, M. Recommended vapor pressures for thiophene, sulfolane, and dimethyl sulfoxide. *Fluid Phase Equilib.* **2011**, *303*, 205–216.

(82) Barley, M. H.; McFiggans, G. The critical assessment of vapour pressure estimation methods for use in modelling the formation of atmospheric organic aerosol. *Atmos. Chem. Phys.* **2010**, *10*, 749–767.

(83) Zielenkiewicz, X.; Perlovich, G. L.; Wszelaka-Rylik, M. The vapour pressure and the enthalpy of sublimation. Determination by inert gas flow method. *J. Therm. Anal. Calorim.* **1999**, *57*, 225–234.

(84) Perlovich, G. L.; Volkova, T. V.; Bauer-Brandl, A. Towards an understanding of the molecular mechanism of solvation of drug molecules: A thermodynamic approach by crystal lattice

energy, sublimation, and solubility exemplified by paracetamol, acetanilide, and phenacetin. *J. Pharm. Sci.* **2006**, *95*, 2158–2169.

(85) Cramer, C. J.; Truhlar, D. G. Am1-Sm2 And Pm3-Sm3 Parameterized Scf Solvation Models For Free-Energies In Aqueous-Solution. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 629–666.

(86) Jorgensen, W. L.; Ulmschneider, J. P.; Tirado-Rives, J. Free energies of hydration from a generalized Born model and an ALL-atom force field. *J. Phys. Chem. B* **2004**, *108*, 16264–16270.

(87) Cabani, S.; Gianni, P.; V., M.; Lepori, L. Group Contributions to the Thermodynamic Properties of Non-ionic Organic Solutes in Dilute Aqueous-solution. *J. Solution Chem.* **1981**, *10*, 563–595.

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

(89) Abraham, M. H.; Andonianhaftvan, J.; Whiting, G. S.; Leo, A.; Taft, R. S. Hydrogen-Bonding .34. The Factors That Influence The Solubility Of Gases And Vapors In Water At 298-K, And A New Method For Its Determination. *J. Chem. Soc., Perkin Trans. 2* **1994**, 1777–1791.

(90) McDonald, N. A.; Carlson, H. A.; Jorgensen, W. L. Free energies of solvation in chloroform and water from a linear response approach. *J. Phys. Org. Chem.* **1997**, *10*, 563–576.

(91) Duffy, E. M.; Jorgensen, W. L. Prediction of properties from simulations: Free energies of solvation in hexadecane, octanol, and water. *J. Am. Chem. Soc.* **2000**, *122*, 2878–2888.

(92) Ratkova, E. L.; Chuev, G. N.; Sergiievskyi, V. P.; Fedorov, M. V. An Accurate Prediction of Hydration Free Energies by Combination of Molecular Integral Equations Theory with Structural Descriptors. *J. Phys. Chem. B* **2010**, *114*, 12068–12079.

(93) Chamberlin, A. C.; Cramer, C. J.; Truhlar, D. G. Predicting Aqueous Free Energies of Solvation as Functions of Temperature. *J. Phys. Chem. B* **2006**, *110*, 5665–5675.

(94) Chamberlin, A. C.; Cramer, C. J.; Truhlar, D. G. Extension of a Temperature-Dependent Aqueous Solvation Model to Compounds Containing Nitrogen, Fluorine, Chlorine, Bromine, and Sulfur. *J. Phys. Chem. B* **2008**, *112*, 3024–3039.

(95) Geballe, M. T.; Guthrie, J. P. The SAMPL3 blind prediction challenge: transfer energy overview. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 489–496.

(96) Mobley, D. L.; Wymer, K. L.; Lim, N. M.; Guthrie, J. P. Blind prediction of solvation free energies from the SAMPL4 challenge. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 135–150.

(97) From private communications with DSc German L. Perlovich, head of the laboratory "Physical Chemistry of Drugs", Institute of Solution Chemistry of Russian Academy of Science.

(98) Garrido, N. M.; Queimada, A. J.; Jorge, M.; Macedo, E. A.; Economou, I. G. 1-Octanol/Water Partition Coefficients of n-Alkanes from Molecular Simulations of Absolute Solvation Free Energies. *J. Chem. Theory Comput.* **2009**, *5*, 2436–2446.

(99) Perlovich, G.; Bauer-Brandl, A. Solvation of Drugs as a Key for Understanding Partitioning and Passive Transport Exemplified by NSAIDs. *Curr. Drug Deliv.* **2004**, *1*, 213–226.

(100) Swanson, J. M. J.; Henchman, R. H.; McCammon, J. A. Revisiting free energy calculations: A theoretical connection to MM/PBSA and direct calculation of the association free energy. *Biophys. J.* **2004**, *86*, 67–74.

(101) Genheden, S.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Ryde, U. An MM/3D-RISM Approach for Ligand Binding Affinities. *J. Phys. Chem. B* **2010**, *114*, 8505–8516.

(102) Palmer, D. S.; Sørensen, J.; Schiøtt, B.; Fedorov, M. V. Solvent Binding Analysis and Computational Alanine Scanning of the Bovine Chymosin–Bovine κ-Casein Complex using Molecular Integral Equation Theory. *J. Chem. Theory Comput.* **2013**, *9*, 5706–5717.

(103) Jensen, J. H.; Li, H.; Robertson, A. D.; Molina, P. A. Prediction and rationalization of protein pK(a) values using QM and QM/MM methods. *J. Phys. Chem. A* **2005**, *109*, 6634–6643.

(104) Jorgensen, W. L.; Duffy, E. M. Prediction of drug solubility from structure. *Adv. Drug Delivery Rev.* **2002**, *54*, 355–366.

(105) Kollman, P. Free-Energy Calculations - Applications To Chemical And Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.

(106) Jorgensen, W. L.; Tirado-Rives, J. Free energies of hydration for organic molecules from Monte Carlo simulations. *Perspect. Drug Discovery Des.* **1995**, *3*, 123–138.

(107) Roux, B.; Simonson, T. Implicit solvent models. *Biophys. Chem.* **1999**, *78*, 1–20.

(108) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889–897.

(109) Orozco, M.; Luque, F. J. Theoretical methods for the description of the solvent effect in biomolecular systems. *Chem. Rev.* **2000**, *100*, 4187–4225.

(110) Westergren, J.; Lindfors, L.; Hoglund, T.; Luder, K.; Nordholm, S.; Kjellander, R. In silico prediction of drug solubility: 1. Free energy of hydration. *J. Phys. Chem. B* **2007**, *111*, 1872–1882.

(111) Nicholls, A.; Mobley, D. L.; Guthrie, J. P.; Chodera, J. D.; Bayly, C. I.; Cooper, M. D.; Pande, V. S. Predicting small-molecule solvation free energies: An informal blind test for computational chemistry. *J. Med. Chem.* **2008**, *51*, 769–779.

(112) Klamt, A.; Eckert, F.; Diedenhofen, M. Prediction of the Free Energy of Hydration of a Challenging Set of Pesticide-Like Compounds. *J. Phys. Chem. B* **2009**, *113*, 4508–4510.

123

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

(114) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Dill, K. A. Predictions of Hydration Free Energies from All-Atom Molecular Dynamics Simulations. *J. Phys. Chem. B* **2009**, *113*, 4533–4537.

(115) Sulea, T.; Wanapun, D.; Dennis, S.; Purisima, E. O. Prediction of SAMPL-1 Hydration Free Energies Using a Continuum Electrostatics-Dispersion Model. *J. Phys. Chem. B* **2009**, *113*, 4511–4520.

(116) van de Waterbeemd, H.; Gifford, E. ADMET in silico modelling: Towards prediction paradise? *Nat. Rev. Drug Discovery* **2003**, *2*, 192–204.

(117) Wells, P., Ed. *Linear free energy relationships*; Academic Press: London, 1968.

(118) Chapman, N. B., Shorter, J., Eds.; *Advances in linear free energy relationships*; Plenum Press: London, 1972.

(119) Shorter, J., Ed. *Correlation Analysis in Organic Chemistry - An Introduction to Linear Free Energy Relationships*; Clarendon Pres: Oxford, 1973.

(120) Voet, D.; Voet, J. G. *Biochemistry*, 3rd ed.; Wiley and Sons, Inc.: Hoboken, NJ, 2002; p 1022.

(121) Michielan, L.; Bacilieri, M.; Kaseda, C.; Moro, S. Prediction of the aqueous solvation free energy of organic compounds by using autocorrelation of molecular electrostatic potential surface properties combined with response surface analysis. *Bioorg. Med. Chem.* **2008**, *16*, 5733–5742.

(122) Tomasi, J.; Persico, M. Molecular-Interactions In Solution - An Overview Of Methods Based On Continuous Distributions Of The Solvent. *Chem. Rev.* **1994**, *94*, 2027–2094.

(123) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **2005**, *105*, 2999–3094.

(124) Cramer, C. J.; Truhlar, D. G. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem. Rev.* **1999**, *99*, 2161–2200.

(125) Reynolds, C. A.; King, P. M.; Richards, W. G. Free-Energy Calculations In Molecular Biophysics. *Mol. Phys.* **1992**, *76*, 251–275.

(126) Matubayasi, N.; Nakahara, M. Theory of solutions in the energetic representation. I. Formulation. *J. Chem. Phys.* **2000**, *113*, 6070–6081.

(127) Matubayasi, N.; Nakahara, M. An approach to the solvation free energy in terms of the distribution functions of the solute-solvent interaction energy. *J. Mol. Liq.* **2005**, *119*, 23–29.

(128) Shirts, M. R.; Pande, V. S. Solvation free energies of amino acid side chain analogs for common molecular mechanics water models. *J. Chem. Phys.* **2005**, *122*, 134508(1-13).

(129) Matubayasi, N. Free-Energy Analysis of Solvation with the Method of Energy Representation. *Front. Biosci.* **2009**, *14*, 3536–3549.

(130) Knight, J. L.; Brooks, C. L. lambda-Dynamics Free Energy Simulation Methods. *J. Comput. Chem.* **2009**, *30*, 1692–1700.

(131) Frolov, A. I.; Kiselev, M. G. Prediction of Cosolvent Effect on Solvation Free Energies and Solubilities of Organic Compounds in Supercritical Carbon Dioxide Based on Fully Atomistic Molecular Simulations. *J. Phys. Chem. B* **2014**, *118*, 11769–11780.

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

(133) Hirata, F. Chemical processes in solution studied by an integral equation theory of molecular liquids. *Bull. Chem. Soc. Jpn.* **1998**, *71*, 1483–1499.

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

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

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

(137) Chuev, G. N.; Fedorov, M. V.; Chiodo, S.; Russo, N.; Sicilia, E. Hydration of ionic species studied by the reference interaction site model with a repulsive bridge correction. *J. Comput. Chem.* **2008**, *29*, 2406–2415.

(138) Luchko, T.; Joung, I. S.; Case, D. A. Integral Equation Theory of Biomolecules and Electrolytes. In *Innovations in Biomolecular Modeling and Simulations: Volume 1*; Schlick, T., Ed.; RSC Biomolecular Sciences; The Royal Society of Chemistry, 2012; Vol. 1; pp 51–86.

(139) Perkyns, J. S.; Lynch, G. C.; Howard, J. J.; Pettitt, B. M. Protein solvation from theory and simulation: Exact treatment of Coulomb interactions in three-dimensional theories. *J. Chem. Phys.* **2010**, *132*, 064106(1-13).

(140) Ten-no, S.; Jung, J.; Chuman, H.; Kawashima, Y. Assessment of free energy expressions in RISM integral equation theory: theoretical prediction of partition coefficients revisited. *Mol. Phys.* **2010**, *108*, 327–332.

(141) Palmer, D. S.; Sergiievskyi, V. P.; Jensen, F.; Fedorov, M. V. Accurate calculations of the hydration free energies of druglike molecules using the reference interaction site model. *J. Chem. Phys.* **2010**, *133*, 044104(1-11).

(142) Monson, P. A.; Morriss, G. P. Recent Progress in the Statistical-mechanics of Interaction Site Fluids. *Adv. Chem. Phys.* **1990**, *77*, 451–550.

(143) Hirata, F., Ed. *Molecular theory of solvation*; Kluwer Academic Publishers, Dordrecht, Netherlands, 2003.

(144) Barrat, J.-L.; Hansen, J.-P. *Basic concepts for simple and complex liquids*; Cambridge University Press, New York, 2003.

(145) Zhou, S.; Solana, J. R. Constructing a new closure theory based on the third-order Ornstein-Zernike equation and a study of the adsorption of simple fluids. *Chem. Rev.* **2009**, *109*, 2829–2858.

(146) Lee, L. L. Constructing a new closure theory based on the third-order Ornstein-Zernike equation and a study of the adsorption of simple fluids. *J. Chem. Phys.* **2011**, *135*, 204706(1–11).

(147) Ornstein, L. S.; Zernike, F. Acculental deviations of density and opalescence at the critical point of a simple substance. *Proceedings of the Koninklijke Akademie Van Wetenschappen Te Amsterdam* **1914**, *17*, 793–806.

(148) Hansen, J.-P.; McDonald, I. R. *Theory of Simple Liquids, 4th ed*; Elsevier Academic Press, Amsterdam, The Netherlands, 2000.

(149) Percus, J. K.; Yevick, G. J. Analysis of Classical Statistical Mechanics By Means of Collective Coordinates. *Phys. Rev.* **1958**, *110*, 1–13.

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

(151) Martynov, G. A.; Sarkisov, G. N. Exact Equations and the Theory of Liquids .5. *Mol. Phys.* **1983**, *49*, 1495–1504.

(152) Llanorestrepo, M.; Chapman, W. G. Bridge Function And Cavity Correlation-Function

From Simulation - Implications On Closure Relations. *Int. J. Thermophys.* **1995**, *16*, 319–326.

(153) Duh, D. M.; Haymet, A. D. J. Integral-Equation Theory For Uncharged Liquids: The Lennard-Jones Fluid And The Bridge Function. *J. Chem. Phys.* **1995**, *103*, 2625–2633.

(154) Duh, D. M.; Henderson, D. Integral equation theory for Lennard-Jones fluids: The bridge function and applications to pure fluids and mixtures. *J. Chem. Phys.* **1996**, *104*, 6742–6754.

(155) Weeks, J. D.; Chandler, D.; Andersen, H. C. Role Of Repulsive Forces In Determining Equilibrium Structure Of Simple Liquids. *J. Chem. Phys.* **1971**, *54*, 5237–5247.

(156) Martynov, G. A.; Sarkisov, G. N.; Vompe, A. G. New closure for the Ornstein-Zernike equation. *J. Chem. Phys.* **1999**, *110*, 3961–3969.

(157) Rogers, F. J.; Young, D. A. New, Thermodynamically Consistent, Integral-equation For Simple Fluids. *Phys. Rev. A* **1984**, *30*, 999–1007.

(158) Zerah, G.; Hansen, J. P. Self-consistent Integral-equations For Fluid Pair Distribution-functions - Another Attempt. *J. Chem. Phys.* **1986**, *84*, 2336–2343.

(159) Verlet, L. Integral-equations For Classical Fluids .2. Hard-spheres Again. *Mol. Phys.* **1981**, *42*, 1291–1302.

(160) Rosenfeld, Y.; Ashcroft, N. W. Mean-spherical Model For Soft Potentials - Hard-core Revealed As A Perturbation. *Phys. Rev. A* **1979**, *20*, 2162–2169.

(161) Malijevsky, A.; Labik, S. The Bridge Function For Hard-spheres. *Mol. Phys.* **1987**, *60*, 663–669.

(162) Yoshida, N.; Kato, S. Molecular Ornstein-Zernike approach to the solvent effects on solute electronic structures in solution. *J. Chem. Phys.* **2000**, *113*, 4974–4984.

(163) Gendre, L.; Ramirez, R.; Borgis, D. Classical density functional theory of solvation in molecular solvents: Angular grid implementation. *Chem. Phys. Lett.* **2009**, *474*, 366–370.

(164) Zhao, S. L.; Ramirez, R.; Vuilleumier, R.; Borgis, D. Molecular density functional theory of solvation: From polar solvents to water. *J. Chem. Phys.* **2011**, *134*, 194102(1-13).

(165) Ishizuka, R.; Yoshida, N. Application of efficient algorithm for solving six-dimensional molecular Ornstein-Zernike equation. *J. Chem. Phys.* **2012**, *136*, 114106(1-6).

(166) Ishizuka, R.; Yoshida, N. Extended molecular Ornstein-Zernike integral equation for fully anisotropic solute molecules: Formulation in a rectangular coordinate system. *J. Chem. Phys.* **2013**, *139*, 084119(1-10).

(167) Chandler, D.; Andersen, H. C. Optimized cluster expansions for classical fluids. 2. Theory of molecular liquids. *J. Chem. Phys.* **1972**, *57*, 1930–1937.

(168) Casanova, D.; Gusarov, S.; Kovalenko, A.; Ziegler, T. 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.* **2007**, *3*, 458–476.

(169) Phongphanphanee, S.; Yoshida, N.; Hirata, F. On the proton exclusion of aquaporins: A statistical mechanics study. *J. Am. Chem. Soc.* **2008**, *130*, 1540–1541.

(170) Frach, R.; Kast, S. M. Solvation Effects on Chemical Shifts by Embedded Cluster Integral Equation Theory. *J. Phys. Chem. A* **2014**, *118*, 11620–11628.

(171) Yamazaki, T.; Kovalenko, A.; Murashov, V. V.; Patey, G. N. Ion Solvation in a Water-Urea Mixture. *J. Phys. Chem. B* **2010**, *114*, 613–619.

(172) Fafard, J.; Lyubimova, O.; Stoyanov, S. R.; Dedzo, G. K.; Gusarov, S.; Kovalenko, A.; Detellier, C. Adsorption of Indole on Kaolinite in Nonaqueous Media: Organoclay Preparation and Characterization, and 3D-RISM-KH Molecular Theory of Solvation Investigation. *J. Phys. Chem. C* **2013**, *117*, 18556–18566.

(173) Heil, J.; Tomazic, D.; Egbers, S.; Kast, S. M. Acidity in DMSO from the embedded cluster integral equation quantum solvation model. *J. Mol. Model.* **2014**, *20*, 1–8.

(174) Kinoshita, M.; Hirata, F. Application of the reference interaction site model theory to analysis on surface-induced structure of water. *J. Chem. Phys.* **1996**, *104*, 8807–8815.

(175) Kinoshita, M.; Hirata, F. Analysis of salt effects on solubility of noble gases in water using the reference interaction site model theory. *J. Chem. Phys.* **1997**, *106*, 5202–5215.

(176) Kinoshita, M.; Okamoto, Y.; Hirata, F. Solvation structure and stability of peptides in aqueous solutions analyzed by the reference interaction site model theory. *J. Chem. Phys.* **1997**, *107*, 1586–1599.

(177) Kinoshita, M.; Okamoto, Y.; Hirata, F. First-principle determination of peptide conformations in solvents: Combination of Monte Carlo simulated annealing and RISM theory. *J. Am. Chem. Soc.* **1998**, *120*, 1855–1863.

(178) Kinoshita, M.; Okamoto, Y.; Hirata, F. Analysis on conformational stability of C-peptide of ribonuclease A in water using the reference interaction site model theory and Monte Carlo simulated annealing. *J. Chem. Phys.* **1999**, *110*, 4090–4100.

(179) Kovalenko, A.; Hirata, F.; Kinoshita, M. Hydration structure and stability of Met-enkephalin studied by a three-dimensional reference interaction site model with a repulsive bridge correction and a thermodynamic perturbation method. *J. Chem. Phys.* **2000**, *113*, 9830–9836.

(180) Kinoshita, M.; Okamoto, Y.; Hirata, F. Solvent effects on conformational stability of peptides: RISM analyses. *J. Mol. Liq.* **2001**, *90*, 195–204.

(181) Kinoshita, M.; Sugai, Y. A new methodology for predicting shape and size distribution of micelles. *Chem. Phys. Lett.* **1999**, *313*, 685–692.

(182) Kinoshita, M.; Sugai, Y. Methodology for predicting approximate shape and size distribution of micelles. *Stud. Surf. Sci. Catal.* **2001**, *132*, 109–112.

(183) Kinoshita, M.; Sugai, Y. Methodology of predicting approximate shapes and size distribution of micelles: illustration for simple models. *J. Comput. Chem.* **2002**, *23*, 1445–1455.

(184) Kinoshita, M.; Okamoto, Y.; Hirata, F. Singular Behavior of the RISM Theory Observed for Peptide in Salt Solution. *Chem. Phys. Lett.* **1998**, *297*, 433–438.

(185) Kinoshita, M.; Okamoto, Y.; Hirata, F. Peptide conformations in alcohol and water: Analyses by the reference interaction site model theory. *J. Am. Chem. Soc.* **2000**, *122*, 2773–2779.

(186) Akiyama, Y.; Kinoshita, M.; Hirata, F. Free energy profiles of electron transfer at water-electrode interface studied by the reference interaction site model theory. *Chem. Phys. Lett.* **1999**, *305*, 251–257.

(187) Imai, T.; Nomura, H.; Kinoshita, M.; Hirata, F. Partial Molar Volume and Compressibility of Alkali-Halide Ions in Aqueous Solution: Hydration Shell Analysis with an Integral Equation Theory of Molecular Liquids. *J. Phys. Chem. B* **2002**, *106*, 7308–7314.

(188) Imai, T.; Kinoshita, M.; Hirata, F. Theoretical study for partial molar volume of amino acids in aqueous solution: Implication of ideal fluctuation volume. *J. Chem. Phys.* **2000**, *112*, 9469–9478.

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

(190) Kinoshita, M.; Imai, T.; Kovalenko, A.; Hirata, F. Improvement of the reference interaction site model theory for calculating the partial molar volume of amino acids and polypeptides. *Chem. Phys. Lett.* **2001**, *348*, 337–342.

(191) Imai, T.; Harano, Y.; Kinoshita, M.; Kovalenko, A.; Hirata, F. A theoretical analysis on hydration thermodynamics of proteins. *J. Chem. Phys.* **2006**, *125*, 024911(1–7).

(192) Imai, T.; Harano, Y.; Kinoshita, M.; Kovalenko, A.; Hirata, F. Theoretical analysis on changes in thermodynamic quantities upon protein folding: Essential role of hydration. *J. Chem. Phys.* **2007**, *126*, 225102(1–9).

(193) Sindhikara, D. J.; Yoshida, N.; Hirata, F. Placevent: An algorithm for prediction of explicit solvent atom distribution-Application to HIV-1 protease and F-ATP synthase. *J. Comput. Chem.* **2012**, *33*, 1536–1543.

(194) Silveira, R. L.; Stoyanov, S. R.; Gusarov, S.; Skaf, M. S.; Kovalenko, A. Plant Biomass Recalcitrance: Effect of Hemicellulose Composition on Nanoscale Forces that Control Cell Wall Strength. *J. Am. Chem. Soc.* **2013**, *135*, 19048–19051.

(195) Silveira, R. L.; Stoyanov, S. R.; Gusarov, S.; Skaf, M. S.; Kovalenko, A. Supramolecular Interactions in Secondary Plant Cell Walls: Effect of Lignin Chemical Composition Revealed with the Molecular Theory of Solvation. *J. Phys. Chem. Lett.* **2015**, *6*, 206–211.

(196) Giambaşu, G. M.; Luchko, T.; Herschlag, D.; York, D. M.; Case, D. A. Ion Counting from Explicit-Solvent Simulations and 3D-RISM. *Biophys. J.* **2014**, *106*, 883–894.

(197) Nikolić, D.; Blinov, N.; Wishart, D.; Kovalenko, A. 3D-RISM-Dock: A New Fragment-Based Drug Design Protocol. *J. Chem. Theory Comput.* **2012**, *8*, 3356–3372.

(198) Bruzzone, S.; Malvaldi, M.; Chiappe, C. A RISM approach to the liquid structure and solvation properties of ionic liquids. *Phys. Chem. Chem. Phys.* **2007**, *9*, 5576–5581.

(199) Bruzzone, S.; Malvaldi, M.; Chiappe, C. Solvation thermodynamics of alkali and halide ions in ionic liquids through integral equations. *J. Chem. Phys.* **2008**, *129*, 074509(1-9).

(200) Yamazaki, T.; Fenniri, H. Imaging Carbon Nanotube Interaction with Nucleobases in Water Using the Statistical Mechanical Theory of Molecular Liquids. *J. Phys. Chem. C* **2012**, *116*, 15087–15092.

(201) Sergiievskyi, V. P.; Hackbusch, W.; Fedorov, M. V. Multigrid Solver for the Reference Interaction Site Model of Molecular Liquids Theory. *J. Comput. Chem.* **2011**, *32*, 1982–1992.

(202) Palmer, D. S.; Chuev, G. N.; Ratkova, E. L.; Fedorov, M. V. In silico screening of bioactive and biomimetic solutes by Integral Equation Theory. *Curr. Pharm. Des.* **2011**, *17*, 1695–1708.

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

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

(205) Kovalenko, A.; Hirata, F. Potential of mean force between two molecular ions in a polar molecular solvent: A study by the three-dimensional reference interaction site model. *J. Phys. Chem. B* **1999**, *103*, 7942–7957.

(206) Kinoshita, M.; Harada, M. Numerical Solution Of The RHNC Theory For Water-Like Fluids Near A Macroparticle And A Planar Wall. *Mol. Phys.* **1994**, *81*, 1473–1488.

(207) Kast, S. M.; Kloss, T. Closed-form expressions of the chemical potential for integral equation closures with certain bridge functions. *J. Chem. Phys.* **2008**, *129*, 236101(1-3).

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

(209) Joung, I. S.; Luchko, T.; Case, D. A. Simple electrolyte solutions: Comparison of RISM and molecular dynamics results for alkali halide solutions. *J. Chem. Phys.* **2013**, *138*, 044103(1-15).

(210) Pettitt, B. M.; Rossky, P. J. Integral-Equation predictions of liquid-state structure for water-like intermolecular potentials. *J. Chem. Phys.* **1982**, *77*, 1451–1457.

(211) Hirata, F.; Pettitt, B. M.; Rossky, P. J. Application of an Extended RISM equation to dipolar and quadrupolar fluids. *J. Chem. Phys.* **1982**, *77*, 509–520.

(212) Hirata, F.; Rossky, P. J.; Pettitt, B. M. The Interionic Potential of Mean Force in a Molecular Polar Solvent from an Extended RISM Equation. *J. Chem. Phys.* **1983**, *78*, 4133–4144.

(213) Pettitt, B. M.; Rossky, J. P. Alkali-Halides in Water: Ion Solvent Correlations and Ion-Ion Potentials of Mean Force At Infinite Dilution. *J. Chem. Phys.* **1986**, *84*, 5836–5834.

(214) Cummings, P. T.; Stell, G. Exact Asymptotic Form of the Site-site Direct Correlation-function for Rigid Polar-molecules. *Mol. Phys.* **1981**, *44*, 529–531.

(215) Perkyns, J.; Pettitt, B. M. Integral-equation Approaches To Structure and Thermodynamics of Aqueous Salt-solutions. *Biophys. Chem.* **1994**, *51*, 129–146.

(216) Yoshida, K.; Yamaguchi, T.; Kovalenko, A.; Hirata, F. Structure of tert-butyl alcohol-water mixtures studied by the RISM theory. *J. Phys. Chem. B* **2002**, *106*, 5042–5049.

(217) Yu, H.-A.; Karplus, M. A thermodynamic analysis of solvation. *J. Chem. Phys.* **1988**, *89*, 2366–2379.

(218) Chong, S. H.; Hirata, F. Ion hydration: Thermodynamic and structural analysis with an integral equation theory of liquids. *J. Phys. Chem. B* **1997**, *101*, 3209–3220.

(219) Chuev, G. N.; Chiodo, S.; Erofeeva, S. E.; Fedorov, M. V.; Russo, N.; Sicilia, E. A quasi-linear RISM approach for the computation of solvation free energy of ionic species. *Chem. Phys. Lett.* **2006**, *418*, 485–489.

(220) Chiodo, S.; Chuev, G. N.; Erofeeva, S. E.; Fedorov, M. V.; Russo, N.; Sicilia, E. Comparative Study of Electrostatic Solvent Response by Rism and Pcm Methods. *Int. J. Quantum Chem.* **2006**, *107*, 265–274.

(221) Fedorov, M. V.; Kornyshev, A. A. Unravelling the solvent response to neutral and charged solutes. *Mol. Phys.* **2007**, *105*, 1–16.

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

(223) Perkyns, J.; Pettitt, B. M. A Site Site Theory For Finite Concentration Saline Solutions. *J. Chem. Phys.* **1992**, *97*, 7656–7666.

(224) Yoshida, N.; Phongphanphanee, S.; Hirata, F. Selective Ion Binding by Protein Probed with the Statistical Mechanical Integral Equation Theory. *J. Phys. Chem. B* **2007**, *111*, 4588–4595.

(225) Hummer, G.; Soumpasis, D. M. An extended RISM study of simple electrolytes-pair correlations in a NaCl-SPC water model. *Mol. Phys.* **1992**, *75*, 633–651.

(226) Hummer, G.; Soumpasis, D. M.; Neumann, M. Pair correlations in an NaCl-SPC water model simulations versus extended RISM computations. *Mol. Phys.* **1992**, *77*, 769–785.

(227) Perkyns, J.; Pettitt, B. M. On the Solubility of Aqueous Electrolytes. *J. Phys. Chem.* **1994**, *98*, 5147-5151.

(228) Fedotova, M. V.; Oparin, R. D.; Trostin, V. N. Structure formation of aqueous electrolyte solutions under extreme conditions by the extended RISM-approach. A possibility of predicting. *J. Mol. Liq.* **2001**, *91*, 123–133.

(229) Fedotova, M. V. Effect of Temperature and Pressure on Structural Self-Organization of Aqueous Sodium Chloride Solutions. *J. Mol. Liq.* **2010**, *153*, 9–14.

(230) Beglov, D.; Roux, B. Solvation of complex molecules in a polar liquid: An integral equation theory. *J. Chem. Phys.* **1996**, *104*, 8678–8689.

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

(232) Beglov, D.; Roux, B. Numerical-Solution Of The Hypernetted-Chain Equation For A Solute Of Arbitrary Geometry In 3 Dimensions. *J. Chem. Phys.* **1995**, *103*, 360–364.

(233) Ratkova, E. L. Computational Prediction of Thermodynamic Properties of Organic Molecules in Aqueous Solutions. Ph.D. Thesis, Fakultät für Chemie der Universität Duisburg-Essen, July 2011.

(234) Miyata, T.; Ikuta, Y.; Hirata, F. Free energy calculation using molecular dynamics simulation combined with the three-dimensional reference interaction site model theory. II. Thermodynamic integration along a spatial reaction coordinate. *J. Chem. Phys.* **2011**, *134*, 044127(1-17).

(235) Kast, S. M. Combinations of simulation and integral equation theory. *Phys. Chem. Chem. Phys.* **2001**, *3*, 5087–5092.

(236) Schmidt, K. F.; Kast, S. M. Hybrid integral equation/Monte Carlo approach to complexation thermodynamics. *J. Phys. Chem. B* **2002**, *106*, 6289–6297.

(237) Tayefeh, S.; Kloss, T.; Thiel, G.; Hertel, B.; Moroni, A.; Kast, S. M. Molecular dynamics simulation of the cytosolic mouth in Kcv-type potassium channels. *Biochemistry* **2007**, *46*, 4826–4839.

(238) Kinoshita, M.; Okamoto, Y.; Hirata, F. Calculation of hydration free energy for a solute with many atomic sites using the RISM theory: A robust and efficient algorithm. *J. Comput. Chem.* **1997**, *18*, 1320–1326.

(239) Kelley, C. T.; Pettitt, B. M. A fast solver for the Ornstein-Zernike equations. *J. Comput. Phys.* **2004**, *197*, 491–501.

(240) Homeier, H.; Rast, S.; Krienke, H. Iterative solution of the Ornstein-Zernike equation with various closures using vector extrapolation. *Comput. Phys. Commun.* **1995**, *92*, 188–202.

(241) Pulay, P. Convergence Acceleration of Iterative Sequences - the Case of Scf Iteration. *Chem. Phys. Lett.* **1980**, *73*, 393–398.

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

(243) Gusarov, S.; Pujari, B. S.; Kovalenko, A. Efficient treatment of solvation shells in 3D molecular theory of solvation. *J. Comput. Chem.* **2012**, *33*, 1478–1494.

(244) Gillan, M. J. New Method of Solving the Liquid Structure Integral-equations. *Mol. Phys.* **1979**, *38*, 1781–1794.

(245) Labik, S.; Malijevsky, A.; Vonka, P. A Rapidly Convergent Method of Solving the Ornstein-Zernike Equation. *Mol. Phys.* **1985**, *56*, 709–715.

(246) Labik, S.; Pospiil, R.; Malijevsky, A.; Smith, W. R. An Efficient Gauss-Newton-like Method for the Numerical Solution of the Ornstein-Zernike Integral Equation for a Class of Fluid Models. *J. Comput. Phys.* **1994**, *115*, 12–21.

(247) Woelki, S.; Kohler, H.; Krienke, H.; Schmeer, G. Improvements of DRISM calculations: symmetry reduction and hybrid algorithms. *Phys. Chem. Chem. Phys.* **2008**, *10*, 898–910.

(248) Chuev, G. N.; Fedorov, M. V. Wavelet algorithm for solving integral equations of molecular liquids. A test for the reference interaction site model. *J. Comput. Chem.* **2004**, *25*, 1369–1377.

(249) Chuev, G. N.; Fedorov, M. V. Wavelet treatment of structure and thermodynamics of simple liquids. *J. Chem. Phys.* **2004**, *120*, 1191–1196.

(250) Kelley, C. T. *Iterative methods for linear and nonlinear equations*; Frontiers in Applied Mathematics; SIAM, Philadelphia, 1995; Vol. 16.

(251) Booth, M. J.; Schlijper, A.; Scales, L.; Haymet, A. Efficient solution of liquid state integral equations using the Newton–GMRES algorithm. *Comput. Phys. Commun.* **1999**, *119*, 122–134.

(252) Hackbusch, W. *Multi-grid methods and Applications*; Springer-Verlag, Berlin, 1985.

(253) Briggs, W. L. *A Multigrid Tutorial*; SIAM, Philadelphia, 1987.

(254) Bank, R.; Dupont, T.; Yserentant, H. The hierarchical basis multigrid methods. *Numerische Mathematik*

(255) Xu, J. Iterative methods by space decomposition and subspace correction. *SIAM Rev.* **1992**, *34*, 581–613.

(256) Vanek, P.; Mandel, J.; Brezina, M. Algebraic multigrid by smoothed aggregation for second and fourth order elliptic problems. *Computing* **1996**, *56*, 179–196, International GAMM-Workshop on Multi-Level Methods, Meisdorf Harz, Germany Sep. 26-28, 1994.

(257) Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. Frontier bonds in QM/MM methods: A comparison of different approaches. *J. Phys. Chem. A* **2000**, *104*, 1720–1735.

(258) Neria, E.; Fischer, S.; Karplus, M. Simulation of activation free energies in molecular systems. *J. Chem. Phys.* **1996**, *105*, 1902–1921.

(259) Karplus, M.; McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **2002**, *9*, 646–652.

(260) Warshel, A.; Russell, S. T. Calculations of Electrostatic Interactions in Biological Systems and in Solutions. *Q. Rev. Biophys.* **1984**, *17*, 283–422.

(261) Shao, Y. et al. Advances in methods and algorithms in a modern quantum chemistry program package. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172–3191.

(262) Brooks, B. R. et al. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.

(263) Field, M. J.; Bash, P. A.; Karplus, M. A Combined Quantum-Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations. *J. Comput. Chem.* **1990**, *11*, 700–733.

(264) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.

(265) MacKerell, A. D. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(266) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.

(267) Gao, J. L. Hybrid quantum and molecular mechanical simulations: An alternative avenue to solvent effects in organic chemistry. *Acc. Chem. Res.* **1996**, *29*, 298–305.

(268) Lin, H.; Truhlar, D. QM/MM: What Have We Learned, Where are We, and Where Do We Go from Here? *Theor. Chem. Acc.* **2007**, *117*, 185–199.

(269) Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229.

(270) Ten-no, S.; Hirata, F.; Kato, S. A Hybrid Approach For The Solvent Effect On The Electronic-Structure Of A Solute Based On The Rism And Hartree-Fock Equations. *Chem. Phys. Lett.* **1993**, *214*, 391–396.

(271) Sato, H.; Hirata, F.; Kato, S. Analytical energy gradient for the reference interaction site model multiconfigurational self-consistent-field method: Application to 1,2-difluoroethylene in aqueous solution. *J. Chem. Phys.* **1996**, *105*, 1546–1551.

(272) Yokogawa, D.; Sato, H.; Sakaki, S. New generation of the reference interaction site model self-consistent field method: Introduction of spatial electron density distribution to the solvation theory. *J. Chem. Phys.* **2007**, *126*, 244504(1-6).

(273) Gill, P. M. W.; Johnson, B. G.; Pople, J. A. Modeling the Potential of A Charge-distribution. *J. Chem. Phys.* **1992**, *96*, 7178–7179.

(274) Sato, H.; Kovalenko, A.; Hirata, F. Self-consistent field, ab initio molecular orbital and three-dimensional reference interaction site model study for solvation effect on carbon monoxide in aqueous solution. *J. Chem. Phys.* **2000**, *112*, 9463–9468.

(275) Kloss, T.; Heil, J.; Kast, S. M. Quantum Chemistry in Solution by Combining 3D Integral Equation Theory with a Cluster Embedding Approach. *J. Phys. Chem. B* **2008**, *112*, 4337–4343.

(276) Sato, H.; Sakaki, S. Comparison of electronic structure theories for solvated molecules: RISM-SCF versus PCM. *J. Phys. Chem. A* **2004**, *108*, 1629–1634.

(277) Nakai, H. Energy density analysis with Kohn-Sham orbitals. *Chem. Phys. Lett.* **2002**, *363*, 73–79.

(278) Vchirawongkwin, V.; Sato, H.; Sakaki, S. RISM-SCF-SEDD Study on the Symmetry Breaking of Carbonate and Nitrate Anions in Aqueous Solution. *J. Phys. Chem. B* **2010**, *114*, 10513–10519.

(279) Tenno, S.; Hirata, F.; Kato, S. Reference Interaction Site Model Self-Consistent-Field Study For Solvation Effect On Carbonyl-Compounds In Aqueous-Solution. *J. Chem. Phys.* **1994**, *100*, 7443–7453.

(280) Sato, H.; Kobori, Y.; Tero-Kubota, S.; Hirata, F. Theoretical study of electronic and solvent reorganization associated with a charging process of organic compounds. I. Molecular and atomic level description of solvent reorganization. *J. Chem. Phys.* **2003**, *119*, 2753–2760.

140

(281) Chong, S.-H.; Miura, S.-i.; Basu, G.; Hirata, F. Molecular theory for the nonequilibrium free energy profile in electron transfer reaction. *J. Phys. Chem.* **1995**, *99*, 10526–10529.

(282) Malvaldi, M.; Bruzzone, S.; Chiappe, C.; Gusarov, S.; Kovalenko, A. Ab Initio Study of Ionic Liquids by KS-DFT/3D-RISM-KH Theory. *J. Phys. Chem. B* **2009**, *113*, 3536–3542.

(283) Fujishige, S.; Kawashima, Y.; Yoshida, N.; Nakano, H. Three-Dimensional Reference Interaction Site Model Self-Consistent Field Study of the Electronic Structure of [Cr(H2O)(6)](3+) in Aqueous Solution. *J. Phys. Chem. A* **2013**, *117*, 8314–8322.

(284) da Costa, L. M.; Hayaki, S.; Stoyanov, S. R.; Gusarov, S.; Tan, X.; Gray, M. R.; Stryker, J. M.; Tykwinski, R.; Carneiro, J. W. d. M.; Sato, H.; Seidl, P. R.; Kovalenko, A. 3D-RISM-KH molecular theory of solvation and density functional theory investigation of the role of water in the aggregation of model asphaltenes. *Phys. Chem. Chem. Phys.* **2012**, *14*, 3922–3934.

(285) da Costa, L. M.; Stoyanov, S. R.; Gusarov, S.; Tan, X.; Gray, M. R.; Stryker, J. M.; Tykwinski, R.; Carneiro, J. W. d. M.; Seidl, P. R.; Kovalenko, A. Density Functional Theory Investigation of the Contributions of pi-pi Stacking and Hydrogen-Bonding Interactions to the Aggregation of Model Asphaltene Compounds. *Energy & Fuels* **2012**, *26*, 2727–2735.

(286) Moreira da Costa, L.; Stoyanov, S. R.; Gusarov, S.; Seidl, P. R.; Walkimar de M. Carneiro, J.; Kovalenko, A. Computational Study of the Effect of Dispersion Interactions on the Thermochemistry of Aggregation of Fused Polycyclic Aromatic Hydrocarbons as Model Asphaltene Compounds in Solution. *J. Phys. Chem. A* **2014**, *118*, 896–908.

(287) Nikolic, D.; Moffat, K. A.; Farrugia, V. M.; Kobryn, A. E.; Gusarov, S.; Wosnick, J. H.; Kovalenko, A. Multi-scale modeling and synthesis of polyester ionomers. *Phys. Chem. Chem. Phys.* **2013**, *15*, 6128–6138.

(288) Stoyanov, S. R.; Yin, C.-X.; Gray, M. R.; Stryker, J. M.; Gusarov, S.; Kovalenko, A. Density functional theory investigation of the effect of axial coordination and annelation on the

absorption spectroscopy of nickel(II) and vanadyl porphyrins relevant to bitumen and crude oils. *Can. J. Chem.* **2013**, *91*, 872–878.

(289) Khalatur, P. G.; Khokhlov, A. R. Hybrid MC/RISM technique for simulation of polymer solutions: Monte Carlo plus RISM integral equations. *Mol. Phys.* **1998**, *93*, 555–572.

(290) Yoshida, N.; Imai, T.; Phongphanphanee, S.; Kovalenko, A.; Hirata, F. Molecular Recognition in Biomolecules Studied by Statistical-Mechanical Integral-Equation Theory of Liquids. *J. Phys. Chem. B* **2009**, *113*, 873–886.

(291) Chandler, D.; Singh, Y.; Richardson, D. M. Excess Electrons In Simple Fluids .1. General Equilibrium-Theory For Classical Hard-Sphere Solvents. *J. Chem. Phys.* **1984**, *81*, 1975–1982.

(292) Grayce, C. J.; Schweizer, K. S. Solvation Potentials For Macromolecules. *J. Chem. Phys.* **1994**, *100*, 6846–6856.

(293) Khalatur, P. G. Hybrid MC-RISM approach to modeling polymers. *Bull. Russ. Acad. Sci. Phys.* **1995**, *59*, 1447–1451.

(294) Khalatur, P. G.; Zherenkova, L. V.; Khokhlov, A. R. Entropy-driven polymer collapse: Application of the hybrid MC/RISM method to the study of conformational transitions in macromolecules interacting with hard colloidal particles. *Eur. Phys. J. B* **1998**, *5*, 881–897.

(295) Mitsutake, A.; Kinoshita, M.; Okamoto, Y.; Hirata, F. Multicanonical algorithm combined with the RISM theory for simulating peptides in aqueous solution. *Chem. Phys. Lett.* **2000**, *329*, 295–303.

(296) Valiev, M.; Bylaska, E.; Govind, N.; Kowalski, K.; Straatsma, T.; Van Dam, H.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T.; de Jong, W. NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Commun.* **2010**, *181*, 1477–1489.

(297) Chuev, G. N.; Valiev, M.; Fedotova, M. V. Integral Equation Theory of Molecular Solvation Coupled with Quantum Mechanical/Molecular Mechanics Method in NWChem Package. *J. Chem. Theory Comput.* **2012**, *8*, 1246–1254.

(298) Ratkova, E. L.; Fedorov, M. V. On a relationship between molecular polarizability and partial molar volume in water. *J. Chem. Phys.* **2011**, *135*, 244109(1-5).

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

(300) Lepori, L.; Gianni, P. Partial molar volumes of ionic and nonionic organic solutes in water: A simple additivity scheme based on the intrinsic volume approach. *J. Solution Chem.* **2000**, *29*, 405–447.

(301) Chalikian, T. V.; Breslauer, K. J. On volume changes accompanying conformational transitions of biopolymers. *Biopolymers* **1996**, *39*, 619–626.

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

(303) Yu, H. A.; Roux, B.; Karplus, M. Solvation Thermodynamics - An Approach From Analytic Temperature Derivatives. *J. Chem. Phys.* **1990**, *92*, 5020–5032.

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

(305) Ben-Amotz, D.; Raineri, F. O.; Stell, G. Solvation Thermodynamics: Theory and Applications. *J. Phys. Chem. B* **2005**, *109*, 6866–6878.

(306) Frenkel, D.; Smit, B. *Understanding molecular simulation*; Academic Press, 2002; p 672.

(307) Kovalenko, A.; Hirata, F. 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.* **2000**, *112*, 10403–10417.

(308) Ten-no, S.; Iwata, S. On the connection between the reference interaction site model integral equation theory and the partial wave expansion of the molecular Ornstein-Zernike equation. *J. Chem. Phys.* **1999**, *111*, 4865–4868.

(309) Kovalenko, A.; Hirata, F. Potentials of mean force of simple ions in ambient aqueous solution. I. Three-dimensional reference interaction site model approach. *J. Chem. Phys.* **2000**, *112*, 10391–10402.

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

(311) Yamazaki, T.; Kovalenko, A. Spatial Decomposition of Solvation Free Energy Based on the 3D Integral Equation Theory of Molecular Liquid: Application to Miniproteins. *J. Phys. Chem. B* **2011**, *115*, 310–318.

(312) Huang, W.; Dedzo, G. K.; Stoyanov, S. R.; Lyubimova, O.; Gusarov, S.; Singh, S.; Lao, H.; Kovalenko, A.; Detellier, C. Molecule–Surface Recognition between Heterocyclic Aromatic Compounds and Kaolinite in Toluene Investigated by Molecular Theory of Solvation and Thermodynamic and Kinetic Experiments. *J. Phys. Chem. C* **2014**, *118*, 23821–23834.

(313) Palmer, S.; Frolov, A. I.; Ratkova, E. L.; Fedorov, M. V. Towards a universal method to calculate hydration free energies: a 3D reference interaction site model with partial molar volume correction. *J. Phys: Cond. Matt.* **2010**, *22*, 492101(1-9).

(314) Huang, W.; Blinov, N.; Kovalenko, A. Octanol–Water Partition Coefficient from 3D–RISM–KH Molecular Theory of Solvation with Partial Molar Volume Correction. *J. Phys. Chem B* **2015**, just accepted, available online http://dx.doi.org/10.1021/acs.jpcb.5b01291.

(315) Truchon, J.-F.; Pettitt, B. M.; Labute, P. A Cavity Corrected 3D-RISM Functional for Accurate Solvation Free Energies. *J. Chem. Theory Comput.* **2014**, *10*, 934–941.

(316) Chuev, G. N.; Fedorov, M. V. Reference interaction site model study of self-aggregating cyanine dyes. *J. Chem. Phys.* **2009**, *131*, 074503(1-8).

(317) Karino, Y.; Fedorov, M. V.; Matubayasi, N. End-point calculation of solvation free energy of amino-acid analogs by molecular theories of solution. *Chem. Phys. Lett.* **2010**, *496*, 351–355.

(318) Sergiievskyi, V. P.; Jeanmairet, G.; Levesque, M.; Borgis, D. Fast computation of solvation free energies with molecular density functional theory: thermodynamic-ensemble partial molar volume corrections. *J. Phys. Chem. Lett.* **2014**, *5*, 1935–1942.

(319) Draper, N. R.; Smith, H. *Applied Regression Analysis (Wiley Series in Probability and Statistics)*, 3rd ed.; Wiley-Interscience, 1998.

(320) Ashbaugh, H. S.; Kaler, E. W.; Paulaitis, M. E. Hydration and conformational equilibria of simple hydrophobic and amphiphilic solutes. *Biophys. J.* **1998**, *75*, 755–768.

(321) Ashbaugh, H. S.; Kaler, E. W.; Paulaitis, M. E. A "universal" surface area correlation for molecular hydrophobic phenomena. *J. Am. Chem. Soc.* **1999**, *121*, 9243–9244.

(322) Gallicchio, E.; Kubo, M. M.; Levy, R. M. Enthalpy-entropy and cavity decomposition of alkane hydration free energies: Numerical results and implications for theories of hydrophobic solvation. *J. Phys. Chem. B* **2000**, *104*, 6271–6285.

(323) Chuev, G. N.; Sokolov, V. F. Hydration of hydrophobic solutes treated by the fundamental measure approach. *J. Phys. Chem. B* **2006**, *110*, 18496–18503.

(324) Wu, J. Z.; Prausnitz, J. M. Pairwise-additive hydrophobic effect for alkanes in water. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 9512–9515.

(325) Levesque, M.; Vuilleumier, R.; Borgis, D. Scalar fundamental measure theory for hard spheres in three dimensions: Application to hydrophobic solvation. *J. Chem. Phys.* **2012**, *137*, 034115(1-9).

(326) Borgis, D.; Gendre, L.; Ramirez, R. Molecular Density Functional Theory: Application to Solvation and Electron-Transfer Thermodynamics in Polar Solvents. *J. Phys. Chem. B* **2012**, *116*, 2504–2512.

(327) Jeanmairet, G.; Levesque, M.; Vuilleumier, R.; Borgis, D. Molecular Density Functional Theory of Water. *J. Phys. Chem. Lett.* **2013**, *4*, 619–624.

(328) Chong, S.-H.; Ham, S. Thermodynamic-Ensemble Independence of Solvation Free Energy. *J. Chem. Theory Comput.* **2015**, *11*, 378–380.

(329) Palmer, D. S.; Frolov, A. I.; Ratkova, E. L.; Fedorov, M. V. Toward a Universal Model To Calculate the Solvation Thermodynamics of Druglike Molecules: The Importance of New Experimental Databases. *Mol. Pharmaceutics* **2011**, *8*, 1423–1429.

(330) Kubinyi, H. *Pharmacokinetic Optimization in Drug Research*; Verlag Helvetica Chimica Acta, 2007; pp 513.

(331) Ball, P. Water as an active constituent in cell biology. *Chem. Rev.* **2008**, *108*, 74–108 .

(332) Agnieszka, K. B. *Thermodynamics - Interaction Studies - Solids, Liquids and Gases*; InTech, 2011.

(333) Kumar, A.; Zhang, K. Y. J. An Investigation on the Effect of Key Water Molecules on Docking Performance in CSARdock Exercise. *J. Chem. Inf. Model.* **2013**, *53*, 1880–1892 .

(334) Kiyota, Y.; Yoshida, N.; Hirata, F. Affinity of small ligands to myoglobin studied by the 3D-RISM theory. *J. Mol. Liq.* **2011**, *159*, 93–98.

(335) Phongphanphanee, S.; Rungrotmongkol, T.; Yoshida, N.; Hannongbua, S.; Hirata, F. Proton Transport through the Influenza A M2 Channel: Three-Dimensional Reference Interaction Site Model Study. *J. Am. Chem. Soc.* **2010**, *132*, 9782–9788.

(336) Hub, J. S.; de Groot, B. L. Comment on "Molecular Selectivity in Aquaporin Channels Studied by the 3D-RISM Theory". *J. Phys. Chem. B* **2011**, *115*, 8364–8366.

(337) Phongphanphanee, S.; Yoshida, N.; Hirata, F. Reply to "Comment on 'Molecular Selectivity in Aquaporin Channels Studied by the 3D-RISM Theory'". *J. Phys. Chem. B* **2011**, *115*, 8367–8369.

(338) FitzPatrick, M.; Champagne, P.; Cunningham, M. F.; Whitney, R. A. A biorefinery processing perspective: Treatment of lignocellulosic materials for the production of value-added products. *Bioresource Technology* **2010**, *101*, 8915–8922.

(339) Sheldon, R. A. Utilisation of biomass for sustainable fuels and chemicals: Molecules, methods and metrics. *Catalysis Today* **2011**, *167*, 3–13.

(340) Xu, L.; Wang, X.; Wang, X. Effects of Zn2+ Binding on the Structural and Dynamic Properties of Amyloid B Peptide Associated with Alzheimer's Disease: Asp1 or Glu11? *ACS Chem. Neurosci.* **2013**, *4*, 1458–1468.

(341) Panteva, M. T.; Dissanayake, T.; Chen, H.; Radak, B. K.; Kuechler, E. R.; Giambasu, G. M.; Lee, T.-S.; York, D. M. Multiscale methods for computational RNA enzymology. *Methods in enzymology* **2015**, *553*, 335–74.

(342) Kitao, A.; Hirata, F.; Go, N. Effects of Solvent On the Conformation and the Collective Motions of A Protein .3. Free-energy Analysis By the Extended Rism Theory. *J. Phys. Chem.* **1993**, *97*, 10231–10235.

(343) Maruyama, Y.; Matsushita, T.; Ueoka, R.; Hirata, F. Solvent and Salt Effects on Structural Stability of Human Telomere. *J. Phys. Chem. B* **2011**, *115*, 2408–2416.

(344) Yamazaki, T.; Fenniri, H.; Kovalenko, A. Structural Water Drives Self-assembly of Organic Rosette Nanotubes and Holds Host Atoms in the Channel. *ChemPhysChem* **2010**, *11*, 361–367.

(345) Maruyama, Y.; Yoshida, N.; Hirata, F. Revisiting the Salt-Induced Conformational Change of DNA with 3D-RISM Theory. *J. Phys. Chem. B* **2010**, *114*, 6464–6471.

(346) Saenger, W.; Hunter, W. N.; Kennard, O. Dna Conformation Is Determined By Economics In the Hydration of Phosphate Groups. *Nature* **1986**, *324*, 385–388.

(347) Palmer, D. S.; Christensen, A. U.; Sørensen, J.; Celik, L.; Qvist, K. B.; Schiøtt, B. Bovine Chymosin: A Computational Study of Recognition and Binding of Bovine kappa-Casein. *Biochemistry* **2010**, *49*, 2563–2573.

(348) Sørensen, J.; Palmer, D. S.; Qvist, K. B.; Schiøtt, B. Initial Stage of Cheese Production: A Molecular Modeling Study of Bovine and Camel Chymosin Complexed with Peptides from the Chymosin-Sensitive Region of kappa-Casein. *J. Agric. Food. Chem.* **2011**, *59*, 5636–5647.

(349) Sørensen, J.; Palmer, D. S.; Schiøtt, B. Hot-Spot Mapping of the Interactions between Chymosin and Bovine kappa-Casein. *J. Agric. Food. Chem.* **2013**, *61*, 7949–7959.

(350) Phillips, K. L.; Sandler, S. I.; Greene, R. W.; Di Toro, D. M. Quantum Mechanical Predictions of the Henry's Law Constants and Their Temperature Dependence for the 209 Polychlorinated Biphenyl Congeners. *Environ. Sci. Tech.* **2008**, *42*, 8412–8418.

(351) Dearden, J. C. In silico prediction of aqueous solubility. *Expert Opin. Drug Discovery* **2006**, *1*, 31–52.

(352) Palmer, D. S.; Mitchell, J. B. O. Is experimental data quality the limiting factor in predicting the aqueous solubility of druglike molecules? *Mol. Pharmaceutics* **2014**, *11*, 2962–2972 .

(353) Jorgensen, W. L. The many roles of computation in drug discovery. *Science* **2004**, *303*, 1813–1818.

(354) Jin, Z.; Tang, Y. P.; Wu, J. A perturbative density functional theory for square-well fluids. *J. Chem. Phys.* **2011**, *134*, 174702(1–10).

(355) Zhao, S.; Wu, J. An efficient method for accurate evaluation of the site-site direct correlation functions of molecular fluids. *Mol. Phys.* **2011**, *109*, 2553–2564.

(356) Zhao, S.; Jin, Z.; Wu, J. New Theoretical Method for Rapid Prediction of Solvation Free Energy in Water. *J. Phys. Chem. B* **2011**, *115*, 6971–6975.

(357) Wu, J.; Jiang, T.; Jiang, D.-e.; Jin, Z.; Henderson, D. A classical density functional theory for interfacial layering of ionic liquids. *Soft Matter* **2011**, *7*, 11222–11231.

(358) Jiang, D.-e.; Meng, D.; Wu, J. Density functional theory for differential capacitance of planar electric double layers in ionic liquids. *Chem. Phys. Lett.* **2011**, *504*, 153–158.

(359) Liu, Y.; Zhao, S.; Wu, J. A Site Density Functional Theory for Water: Application to Solvation of Amino Acid Side Chains. *J. Chem. Theory Comput.* **2013**, *9*, 1896–1908.

(360) Liu, Y.; Fu, J.; Wu, J. High-Throughput Prediction of the Hydration Free Energies of Small Molecules from a Classical Density Functional Theory. *J. Phys. Chem. Lett.* **2013**, *4*, 3687–3691.

(361) Jiang, D.-e.; Wu, J. Microscopic Insights into the Electrochemical Behavior of Nonaqueous Electrolytes in Electric Double-Layer Capacitors. *J. Phys. Chem. Lett.* **2013**, *4*, 1260–1267.

(362) Henderson, D.; Lamperski, S.; Bhuiyan, L.B.; Wu, J. The tail effect on the shape of an electrical double layer differential capacitance curve. *J. Chem. Phys.* **2013**, *138*, 144704(1–3).

(363) Jiang, D.-e.; Jin, Z.; Henderson, D.; Wu, J. Solvent Effect on the Pore-Size Dependence of an Organic Electrolyte Supercapacitor. *J. Phys. Chem. Lett.* **2012**, *3*, 1727–1731.

(364) Jiang, D.-e.; Jin, Z.; Wu, J. Oscillation of Capacitance i nside Nanopores. *Nano Lett.* **2011**, *11*, 5373–5377.

(365) Jeanmairet, G.; Levesque, M.; Sergiievskyi, V.; Borgis, D. Molecular density functional theory for water with liquid–gas coexistence and correct pressure. *J. Chem. Phys.* **2015**, *142*, 154112(1–7).

(366) Fu, J.; Liu, Y.; Wu, J. Molecular density functional theory for multiscale modeling of hydration free energy. *Chem. Eng. Sci.* **2015**, *126*, 370–382.

(367) Ramirez, R.; Gebauer, R.; Mareschal, M.; Borgis, D. Density functional theory of solvation in a polar solvent: Extracting the functional from homogeneous solvent simulations. *Phys. Rev. E* **2002**, *66*, 031206(1–8).

(368) Ramirez, R.; Mareschal, M.; Borgis, D. Direct correlation functions and the density functional theory of polar solvents. *Chem. Phys.* **2005**, *319*, 261–272.

(369) Chong, S.-H.; Ham, S. Aqueous interaction site integral-equation theory that exactly takes into account intramolecular correlations. *J. Chem. Phys.* **2012**, *137*, 154101(1-17).

(370) Chandler, D.; Mccoy, J. D.; Singer, S. J. Density Functional Theory of Nonuniform Polyatomic Systems. 1. General Formulation. *J. Chem. Phys.* **1986**, *85*, 5971–5976.

(371) Chandler, D.; Mccoy, J. D.; Singer, S. J. Density Functional Theory of Nonuniform Polyatomic Systems. 2. Rational Closures for Integral-Equations. *J. Chem. Phys.* **1986**, *85*, 5977–5982.

(372) Matubayasi, N.; Nakahara, M. Theory of solutions in the energy representation. II. Functional for the chemical potential. *J. Chem. Phys.* **2002**, *117*, 3605–3616.

(373) Matubayasi, N.; Nakahara, M. Theory of solutions in the energy representation. III. Treatment of the molecular flexibility. *J. Chem. Phys.* **2003**, *119*, 9686–9702.

(374) Matubayasi, N.; Nakahara, M. Hydrothermal reactions of formaldehyde and formic acid: Free-energy analysis of equilibrium. *J. Chem. Phys.* **2005**, *122*, 074509(1–12).

(375) Matubayasi, N.; Liang, K.; Nakahara, M. Free-energy analysis of solubilization in micelle. *J. Chem. Phys.* **2006**, *124*, 154908(1-13).

(376) Matubayasi, N.; Shinoda, W.; Nakahara, M. Free-energy analysis of the molecular binding into lipid membrane with the method of energy representation. *J. Chem. Phys.* **2008**, *128*, 195107(1-13).

(377) Saito, H.; Matubayasi, N.; Nishikawa, K.; H., N. Hydration property of globular proteins: An analysis of solvation free energy by energy representation method. *Chem. Phys. Lett.* **2010**, *497*, 218–222.

(378) Karino, Y.; ; Matubayasi, Communication: Free-energy analysis of hydration effect on protein with explicit solvent: Equilibrium fluctuation of cytochrome c. *J. Chem. Phys.* **2011**, *134*, 041105(1-4).

(379) Takahashi, H.; Matubayasi, N.; Nakahara, M.; Nitta, T. A quantum chemical approach to the free energy calculations in condensed systems: The QM/MM method combined with the theory of energy representation. *J. Chem. Phys.* **2004**, *121*, 3989–3999.

(380) Takahashi, H.; Ohno, H.; Yamauchi, T.; Kishi, R.; Furukawa, S.; Nakano, M.; Matubayasi, N. Investigation of the dominant hydration structures among the ionic species in aqueous solution: Novel quantum mechanics/molecular mechanics simulations combined with the theory of energy representation. *J. Chem. Phys.* **2008**, *128*, 064507(1–12).

(381) Takahashi, H.; Ohno, H.; Kishi, R.; Nakano, M.; Matubayasi, N. Computation of the free energy change associated with one-electron reduction of coenzyme immersed in water: A novel approach within the framework of the quantum mechanical/molecular mechanical method combined with the theory of energy representation. *J. Chem. Phys.* **2008**, *129*, 205103(1–14).

(382) Takahashi, H.; Maruyama, K.; Karino, Y.; Morita, A.; Nakano, M.; Jungwirth, P.; Matubayasi, N. Energetic origin of proton affinity to the air/water interface. *J. Phys. Chem. B* **2011**, *115*, 4745–4751.

(383) Matubayasi, N.; Takahashi, H. Free-energy analysis of the electron-density fluctuation in the quantum-mechanical/molecular-mechanical simulation combined with the theory of energy representation. *J. Chem. Phys.* **2012**, *136*, 044505(1–10).

(384) Takeuchi, M., Matubayasi, N., Kameda, Y., Minofar, B., Ishiguro, S., Umebayashi, Y. Free-Energy and Structural Analysis of Ion Solvation and Contact Ion-Pair Formation of Li^+ with BF_4^- and PF_6^- in Water and Carbonate Solvents. *J. Phys. Chem. B* **2012**, *116*, 6476–6487.

(385) Frolov, A. I. Accurate Calculation of Solvation Free Energies in Supercritical Fluids by Fully Atomistic Simulations: Probing the Theory of Solutions in Energy Representation. *J. Chem. Theory Comput.* **2015**, just accepted (available online), DOI: 10.1021/acs.jctc.5b00172 .

(386) Frolov, A. I.; Ratkova, E. L.; Palmer, D. S.; Fedorov, M. V. Supporting Information. Hydration Thermodynamics using the Reference Interaction Site Model: Speed or Accuracy? *J. Phys. Chem. B* **2011**, *115*, 6011–6022, Supporting Information.

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

Ekaterina L. Ratkova is a Junior Researcher at the Russian Academy of Science.

She obtained both her bachelor's (2007) and master's degrees (2009) in Chemistry from the Ivanovo State University of Chemistry and Technology, Russia (both Honors, highest distinction). In 2009 she moved to the Max Planck Institute for Mathematics in the Sciences, Germany, as a research assistant, and started her doctoral studies under the supervision of Prof. DSc. Maxim V. Fedorov. She received her Ph.D. degree (2011) in theoretical chemistry from the University of Duisburg-Essen, Germany (summa cum laude, highest distinction). Her doctoral studies were focused on computational prediction of thermodynamics properties of organic molecules in aqueous solutions. In particular, she was one of the developers of the RISM-SDC computational approach for prediction of solvation free energies which is based on combination of the integral equation theory of molecular liquids and cheminformatics. Several articles resulted from this project are featured in the journals' most-cited publication lists. Since 2012, she has been carrying out computational studies of aqueous solubility of drug-like compounds and pesticides as a junior researcher at the Institute of Solution Chemistry of the Russian Academy of Sciences. She received several prestigious research awards, including the Scholarship of the President of Russian Federation (2013-2015, Russia) [the National award for young researchers]. Her current research interests include computational prediction as well as in vitro measurements of key physical/chemical parameters of drug-candidates: solvation free energy, lipophilicity, protein affinity, etc.

David S. Palmer is a Lecturer in the Department of Chemistry of the University of Strathclyde.

He obtained his PhD in Chemistry from Cambridge University in 2008 for work on computational drug discovery under the supervision of Dr. John Mitchell. This research was carried out in the Unilever Centre for Molecular Informatics and was funded by Pfizer Inc. through the Pfizer Institute for Pharmaceutical Materials. He holds a Masters degree in Chemistry from the University of Sheffield, where he worked on sequence-dependent DNA structure in the group of Prof. Christopher A. Hunter. After completing his doctorate, he carried out postdoctoral research on protein biochemistry with Prof Frank Jensen and Prof Birgit Schiott at Aarhus University in Denmark and on solvation thermodynamics with Prof. DSc Maxim V. Fedorov at the Max Planck Institute for Mathematics in the Sciences in Leipzig, Germany. In 2012, he was awarded a Marie Curie Intra-European Fellowship and moved to the Department of Physics at the University of Strathclyde. He has been a Lecturer in the Department of Chemistry at the same institute since the beginning of 2014. He currently leads a research group applying molecular simulation methods to problems in biochemistry and chemical physics.

Maxim V. Fedorov is an expert in computational physical chemistry, electrochemistry and molecular biophysics. His research interests focus on (i) modeling of molecular-scale effects at solvation interfaces; (ii) ion effects on (bio)macromolecules and nanobjects; (iii) ionic liquids at neutral and charged interfaces. He published over 70 peer-reviewed publications and presented about 70 invited talks and lectures on these subjects.

In 2002 he received his Ph.D. degree in Biophysics at the Institute of Theoretical and Experimental Biophysics of the Russian Academy of Sciences, Pushchino, Russia, with Prof. S.E. Shnoll, an expert in physical biochemistry. After his PhD he worked as a researcher at the same institute until the beginning of 2004 when he moved to the University College Dublin, Republic of Ireland, as a postdoc at the Department of Chemistry. In 2005 he moved to the University of Cambridge, UK, where he spent three years working at the Department of Chemistry as a Research Associate in the Unilever Centre for Molecular Science Informatics. Since 2007 he was a Group Leader at the Max Planck Institute for Mathematics in the Sciences, in Leipzig, Germany. In 2007 he received his DSc (Doktor Nauk) degree in Physical Chemistry from the Institute of Solution Chemistry of the Russian Academy of Sciences, Ivanovo, Russia. In 2010-2011 he was also an external Privat-Dozent at the Department of Chemistry, University of Duisburg-Essen, Germany where he received *Venia Legendi* in 2010. As a visiting scientist he worked at the University of Kassel in 2003 (with Prof. D. Kolb, as a DAAD scholar); Imperial College London in 2007 (with Prof. A.A. Kornyshev); and at the University of Osaka in 2009 (with Prof. Y. Goto). In 2011 he spent a month in the University of Lille as a Visiting Professor in the group of Prof. A. Idrissi. In 2012 he received the Helmholtz Award from the International Association of the Properties of Water and Steam (IAWPS) for his works on theoretical physical chemistry of liquids.

He organized several international conferences, summer schools and workshops on biomolecule solvation, chemical physics and physical chemistry of liquids and applications of high-performance computing in natural sciences and engineering.

In 2011 he joined the University of Strathclyde in Glasgow (UK) as a Professor in Physics and Life Sciences. Since 2012 he is also Director of the West of Scotland Academia-Industry Supercomputer Centre (aka ARCHIE-WeST) at the University of Strathclyde. Since 2012 he is serving as Deputy Leader of Physics and Life Sciences Theme of Scottish Universities Physics Alliance (SUPA).