NRC·CNRC

NRC Publications Archive
Archives des publications du CNRC

Molecular theory of solvation: methodology summary and illustrations

Kovalenko, A.

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.
For the publisher's version, please access the DOI link below./ Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

Publisher's version / Version de l'éditeur:
https://doi.org/10.5488/CMP.18.32601
Condensed Matter Physics, 18, 3, pp. 32601-1-32601-24, 2015-09

NRC Publications Record / Notice d'Archives des publications de CNRC:
https://nrc-publications.canada.ca/eng/view/object/?id=53020f23-1ac2-41af-8879-1819d3a6fe4f
https://publications-cnrc.canada.ca/fra/voir/objet/?id=53020f23-1ac2-41af-8879-1819d3a6fe4f

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at
https://nrc-publications.canada.ca/eng/copyright
READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site
https://publications-cnrc.canada.ca/fra/droits
LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

National Research Council Canada | Conseil national de recherches Canada

Canada

Condensed Matter Physics, 2015, Vol. 18, No 3, 32601: 1–24
DOI: 10.5488/CMP.18.32601
http://www.icmp.lviv.ua/journal

CONDENSED MATTER PHYSICS

Review

Molecular theory of solvation: Methodology summary and illustrations*

A. Kovalenko[1,2]

[1] National Institute for Nanotechnology, 11421 Saskatchewan Dr., Edmonton, AB, T6G 2M9, Canada

[2] Department of Mechanical Engineering, University of Alberta, Mechanical Engineering Bldg. 4–9, Edmonton, AB, T6G 2G7, Canada

Received March 16, 2015, in final form June 22, 2015

Integral equation theory of molecular liquids based on statistical mechanics is quite promising as an essential part of multiscale methodology for chemical and biomolecular nanosystems in solution. Beginning with a molecular interaction potential force field, it uses diagrammatic analysis of the solvation free energy to derive integral equations for correlation functions between molecules in solution in the statistical-mechanical ensemble. The infinite chain of coupled integral equations for many-body correlation functions is reduced to a tractable form for 2- or 3-body correlations by applying the so-called closure relations. Solving these equations produces the solvation structure with accuracy comparable to molecular simulations that have converged but has a critical advantage of readily treating the effects and processes spanning over a large space and slow time scales, by far not feasible for explicit solvent molecular simulations. One of the versions of this formalism, the three-dimensional reference interaction site model (3D-RISM) integral equation complemented with the Kovalenko-Hirata (KH) closure approximation, yields the solvation structure in terms of 3D maps of correlation functions, including density distributions, of solvent interaction sites around a solute (supra)molecule with full consistent account for the effects of chemical functionalities of all species in the solution. The solvation free energy and the subsequent thermodynamics are then obtained at once as a simple integral of the 3D correlation functions by performing thermodynamic integration analytically. Analytical form of the free energy functional permits the self-consistent field coupling of 3D-RISM-KH with quantum chemistry methods in multiscale description of electronic structure in solution, the use of 3D maps of potentials of mean force as scoring functions for molecular recognition and protein-ligand binding in docking protocols for fragment based drug design, and the hybrid MD simulation running quasidynamics of biomolecules steered with 3D-RISM-KH mean solvation forces. The 3D-RISM-KH theory has been validated on both simple and complex associating liquids with different chemical functionalities in a wide range of thermodynamic conditions, at different solid-liquid interfaces, in soft matter, and various environments and confinements. The 3D-RISM-KH theory offers a "mental microscope" capable of providing an insight into structure and molecular mechanisms of formation and functioning of various chemical and biomolecular systems and nanomaterials.

Key words: *solution chemistry, biomolecular solvation, integral equation theory of liquids, 3D-RISM-KH molecular theory of solvation, solvation structure and thermodynamics, potential of mean force*

PACS: *61.20.-p, 61.20.Gy, 65.20.-w, 81.16.Fg, 82.60.-s, 87.15.A-*

1. Introduction

Nanoscale properties, phenomena, and processes are profoundly different from macroscopic laws governing the behavior of continuous media and materials. All functional features of nanostructures stem from the microscopic properties of constituting atoms but manifest at length scale from one to hundreds nanometers and time scale up to microseconds and more. By changing size, composition, and

*This paper is a part of the collection of papers published in Condens. Matter Phys., 2015, **18**, No 1.

© A. Kovalenko, 2015

fabrication protocol, nanostructure properties and processes can be tuned in the widest range [1]. Therefore, predictive modelling of nanosystems should operate at length scales from an Angström to hundreds nanometers and microns and time scales to milliseconds and seconds, and yet derive their properties from the chemical functionalities of the constituents. (An example is biological cellular systems of protein nanomachines operating in crowded environment.) Explicit molecular modelling of such nanosystems involves millions and billions of molecules and is by far not feasible in a "brute force" approach employing just *ab initio* quantum chemical methods and/or molecular simulations. Addressing this challenge requires the development and use of multiscale methods coupling several levels of description, from electronic structure methods for building blocks and classical molecular simulations for critical aggregates in the system, to statistical-mechanical theories for their large assemblies and mean properties in a statistical ensemble over characteristic size and time scales, to eventually come up with macroscopic scale properties of the nanostructures and related processes showing up in the "real observable world". A true, genuine challenge of multiscale modelling is the development of a theoretical framework that couples methods at different scales, so that observables at lower-level scales are analytically linked in a self-consistent field description to force fields of more coarse-grained models at higher-level scales [2]. Statistical mechanics itself, in particular, integral equation theory of liquids [3], is an example of such a theoretical coupling between microscopic molecular variables and thermodynamic, macroscopic properties.

2. Statistical-mechanical, molecular theory of solvation

Molecular theory of solvation, also known as reference interaction site model (RISM) [3, 4], is based on the first principles foundation of statistical mechanics and Ornstein-Zernike (OZ) type integral equation theory of molecular liquids [3]. It provides a firm platform to handle complex chemical and biomolecular systems in solution. As distinct from molecular simulations which explore the phase space of a molecular system by direct sampling, RISM theory operates with spatial distributions rather than trajectories of molecules and is based on analytical summation of the free energy diagrams which yields the solvation structure and thermodynamics in the statistical-mechanical ensemble. It yields a solvation structure by solving the RISM integral equations for correlation functions and then the solvation thermodynamics analytically as a single integral of a closed form in terms of the correlation functions obtained. Its three-dimensional (3D) version, 3D-RISM theory gives 3D maps of solvent distributions around a solute macromolecule of arbitrary shape [5–14]. An important component enabling a utility of 3D-RISM theory for complex systems in solution has been the closure relation proposed by Kovalenko and Hirata (KH approximation) [9, 12, 14]. For simple and complex solvents and solutions of a given composition, including buffers, salts, polymers, ligands and other cofactors at a finite concentration, the 3D-RISM-KH molecular theory of solvation properly accounts for chemical functionalities by representing in a single formalism both electrostatic and non-polar features of solvation, such as hydrogen bonding, structural solvent molecules, salt bridges, solvophobicity, and other electrochemical, associative and steric effects. For real systems, to solve the 3D-RISM-KH integral equations is far less computationally expensive than to run molecular simulations which should be long enough to sample relevant exchange and binding events. This enables handling complex systems and processes occurring on large space and long time scales, which is problematic and frequently not feasible for molecular simulations. The 3D-RISM-KH theory has been validated on both simple and complex associating liquids and solutes with different chemical functionalities [14–25], including ionic liquids [19], and polyelectrolyte gels [20], in a range of fluid thermodynamic conditions [4, 21–24], in various environments such as interfaces with metal [10, 12], metal oxide [26], zeolite [13, 27], clay [28, 29], and in confinement of carbon nanotubes [16], synthetic organic rosette nanotubes [14, 30–32], nanocellulose based bionanomaterials [33–35], and biomolecular systems [14, 36–61]. The latter include case studies ranging from the structure of hydrated alanine dipeptide [36–39], miniprotein 1L2Y and protein G [39]; structural water, xenon, and ions bound to lysozyme protein [40, 41]; selective binding and permeation of water, ions and protons in channels [14, 20, 41–45]; salt-induced conformational transitions of DNA [45–47]; partial molar volume and pressure induced conformational transitions of biomolecules [48–52]; formation and conformational stability of β-sheet Amyloid-β (Aβ) oligomers [53], HET-s Prion and Aβ fibrillous aggregates [54, 55]; ligand-binding affinities

of seven biotin analogues to avidin in aqueous solution [56]; binding modes of inhibitors of pathologic conversion and aggregation of prion proteins [20, 55], binding modes of thiamine against extracytoplasmic thiamine binding lipoprotein MG289 [57, 58], binding modes and ligand efflux pathway in multidrug transporter AcrB [59], and maltotriose ligand binding to maltose-binding protein in which structural water triggers binding modes with protein domain motion resulting in the holo-closed structure that binds the ligand [60]; to aqueous electrolyte solution at physiological concentrations in biomolecular systems as large as the Gloebacter violaceus pentameric ligand-gated ion channel (GLIC) homologue in a lipid bilayer [14, 20, 57] and structural water promoting folding in the GroEL/ES chaperonin complex [61]. The 3D-RISM-KH theory provided an insight into such soft matter phenomena as the structure and stability of gels formed by oligomeric polyelectrolyte gelators in different solvents [20].

2.1. 3D-RISM integral equations for molecular liquid structure

The solvation structure is represented by the probability density $\rho_\gamma g_\gamma(\mathbf{r})$ of finding interaction site γ of solvent molecules at 3D space position $\mathbf{r}$ around the solute molecule (which can be both a macromolecule and supramolecule), as given by the average number density ρ_γ in the solution bulk times the normalized density distribution, or 3D distribution function, $g_\gamma(\mathbf{r})$. The values of $g_\gamma(\mathbf{r}) > 1$ and $g_\gamma(\mathbf{r}) < 1$ indicate the areas of density enhancement or depletion, respectively, relative to the average density at a distance from the solute in the solution bulk where $g_\gamma \to 1$. The 3D distribution functions of solvent interaction sites around the solute molecule are determined from the 3D-RISM integral equation [5–14]

$$h_\gamma(\mathbf{r}) = \sum_\alpha \int \mathrm{d}\mathbf{r}' c_\alpha(\mathbf{r}-\mathbf{r}')\chi_{\alpha\gamma}(r')\,, \tag{2.1}$$

where $h_\gamma(\mathbf{r})$ and $c_\gamma(\mathbf{r})$ are, respectively, the 3D total and direct correlation functions of solvent site γ around the solute molecule; $\chi_{\alpha\gamma}(r)$ is the site-site susceptibility of solvent (comprising both the intra- and intermolecular terms) which is an input to the 3D-RISM theory; and indices α and γ enumerate all interaction sites on all sorts of solvent species. The diagrammatic analysis relates the density distribution function to the total correlation function as [3]

$$g_\gamma(\mathbf{r}) = h_\gamma(\mathbf{r}) + 1\,, \tag{2.2}$$

and so the latter has the meaning of the normalized probability density of 3D spatial correlations between the solute and solvent molecules, or normalized deviations of solvent site density around the solute molecule from its average value in the solution bulk. As follows from the diagrammatic expansion of the direct correlation function [3], the leading term of its asymptotics beyond the short-range region D^{sr} of the solute-solvent repulsive core and attractive well (first solvation shell maximum),

$$c_\gamma(\mathbf{r}) \sim -u_\gamma(\mathbf{r})/(k_\mathrm{B}T) \quad \text{for} \quad \mathbf{r} \notin D^{\mathrm{sr}}\,, \tag{2.3}$$

is given by the 3D interaction potential $u_\gamma(\mathbf{r})$ between the whole solute molecule and solvent interaction site γ, scaled by the Boltzmann factor k_B times temperature T. Inside the repulsive core, the direct correlation function strongly deviates from the asymptotics (2.3) and assumes the values related to the solvation free energy of the solute molecule immersed in the solvent.

2.2. KH closure approximation

To obtain correlation functions, the 3D-RISM integral equation (1) should be complemented with another relation between the two functions $h_\gamma(\mathbf{r})$ and $c_\gamma(\mathbf{r})$ which is called a closure and which also involves the 3D solute-solvent site interaction potential $u_\gamma(\mathbf{r})$ specified with a molecular force field. The exact closure has a nonlocal functional form which can be expressed as an infinite diagrammatic series in terms of multiple integrals of the total correlation function [3]. However, these diagrams are cumbersome and the series suffers from a usual drawback of poor convergence, which makes useless to truncate the series at reasonably low-order diagrams and thus renders the exact closure computationally intractable. Therefore, it is replaced in practice with amenable approximations having analytical features that properly

represent physical characteristics of the system, such as the long-range asymptotics of the correlation functions related to electrostatic forces and their short-range features related to the solvation structure and thermodynamics. Examples of closure relations to Ornstein-Zernike type integral equations (including RISM) suitable for liquids of polar and charged species are the so-called hypernetted chain (HNC) closure and the mean spherical approximation (MSA) [3], since they both enforce the long-range asymptotics like (2.3), which is critical for systems with charges. Such other well-known approximations as the Percus-Yevick, Modified Verlet, Martynov-Sarkisov, and Ballone-Pastore-Galli-Gazzillo closures reproduce solvation features for species with short-range repulsion but do not properly account for electrostatics in the interaction potential. By generalization, the 3D-HNC and 3D-MSA closures to the 3D-RISM integral equation (2.1) are constructed [5–9, 12]. While enforcing the asymptotics (2.3), the 3D-HNC closure strongly overestimates the association effects and, therefore, the 3D-RISM-HNC equations diverge for macromolecules with considerable site charges solvated in polar solvents or electrolyte solutions, which is almost always the case for biomolecules in aqueous solution. The 3D-MSA closure is free from that shortcoming but suffers from another serious deficiency of producing nonphysical negative values of the distribution function in the areas adjacent to the associative peaks. Those drawbacks are overcome with the approximation proposed by Kovalenko and Hirata (KH closure), the 3D version of which reads [9, 12, 14]

$$g_\gamma(\mathbf{r}) = \begin{cases} \exp\left[-u_\gamma(\mathbf{r})/(k_\mathrm{B}T) + h_\gamma(\mathbf{r}) - c_\gamma(\mathbf{r})\right] & \text{for} \quad g_\gamma(\mathbf{r}) \leqslant 1, \\ 1 - u_\gamma(\mathbf{r})/(k_\mathrm{B}T) + h_\gamma(\mathbf{r}) - c_\gamma(\mathbf{r}) & \text{for} \quad g_\gamma(\mathbf{r}) > 1. \end{cases} \tag{2.4}$$

The 3D-KH closure relation (2.4) couples in a nontrivial way the HNC approximation and the MSA linearization, the latter automatically applied to spatial regions of solvent density enhancement $g_\gamma(\mathbf{r}) > 1$, including the repulsive core, and the latter to spatial regions of solvent density enrichment $g_\alpha(\mathbf{r}) > 1$ such as strong peaks of association and long-range tails of near-critical fluid phases, and the former to spatial regions of density depletion $g_\alpha(\mathbf{r}) < 1$ (including the repulsive core), while keeping the right asymptotics (2.3) peculiar in both HNC and MSA. The distribution function and its first derivative are continuous at the joint boundary $g_\alpha(\mathbf{r}) = 1$ by construct. The KH approximation consistently accounts for both electrostatic and non-polar (associative and steric) effects of solvation in simple and complex liquids, non-electrolyte and electrolyte solutions, compounds and supramolecular solutes in various chemical [9–29], soft matter [20], synthetic organic supramolecular [14, 30–32], biopolymeric [33–35], and biomolecular [14, 36–61] systems.

The 3D-KH closure underestimates the height of strong associative peaks of the 3D site distribution functions due to the MSA linearization applied to them [9, 62]. However, it somewhat widens the peaks and so 3D-RISM-KH quite accurately reproduces the coordination numbers of the solvation structure in different systems, including micromicelles in water-alcohol solutions [14, 23–25], solvation shells of metal-water [9, 12], metal oxide-water [26] and mixed organic solvent-clay [28, 29] interfaces, and structural water solvent localized in biomolecular confinement [41, 60, 61]. For instance, the coordination numbers of water strongly bound to the MgO surface are calculated from the 3D-RISM-KH theory with a 90% accuracy and the peak positions within a 0.5 Å deviation, compared to molecular dynamics (MD) simulations [26]. The 3D solvation map $g_\gamma(\mathbf{r})$ of function-related structural water in the GroEL chaperon complex (shown to be strongly correlated to the rate of protein folding inside the chaperon cavity) obtained in an expensive MD simulation with explicit solvent involving 1 million atoms is reproduced from the 3D-RISM-KH theory in a relatively short calculation on a workstation with an accuracy of over 90% correlation for the 3D density map and about 98% correlation for the 3D density maxima [61].

Note that different approximate 3D bridge functions, or bridge corrections, to the 3D-HNC approximation can be constructed, such as the 3D-HNCB closure which reproduced MD simulation results for the height of the 3D distribution peaks of water solvent around the bovine pancreatic trypsin inhibitor (BPTI) protein [62]. A related promising approach is the partial series expansions of order n (PSE-n) of the HNC closure [63]. The PSE-n closures interpolate between the KH and HNC approximations and, therefore, combine numerical stability with enhanced accuracy, as demonstrated for aqueous solutions of alkali halide ions [63, 64]. However, a particular appeal of the 3D-KH closure is that it provides an adequate accuracy and the existence of solutions if expected from physical considerations for a wide class of solution systems “from the wild”, including various chemical and biomolecular solutes in solution with different solvents, co-solvents, and electrolytes [9–61], and moreover, with ligand molecules or fragments consid-

ered as part of solvent for efficient 3D mapping of binding affinities in such problems as molecular recognition and fragment based drug design [41, 58, 59]. Construction of bridge functions for such systems with different solvent components other than water constitutes a significant challenge not addressed so far, and the 3D-KH closure offers a reliable choice for a given system.

2.3. Dielectrically consistent RISM theory for site-site correlations of solvent

The radially dependent site-site susceptibility of bulk solvent breaks up into the intra- and intermolecular terms,

$$\chi_{\alpha\gamma}(r) = \omega_{\alpha\gamma}(r) + \rho_\alpha h_{\alpha\gamma}(r)\,. \tag{2.5}$$

The intramolecular correlation function $\omega_{\alpha\gamma}(r)$ normalized as $4\pi\int r^2 \mathrm{d}r\,\omega_{\alpha\gamma}(r) = 1$ represents the geometry of solvent molecules. [Note that $\omega_{\alpha\gamma}(r) = 0$ for sites α and γ on different species.] For rigid species with site separations $l_{\alpha\gamma}$, it is represented in the direct space in terms of δ-functions as $\omega_{\alpha\gamma}(r) = \delta_{\alpha\gamma}(r - l_{\alpha\gamma})/(4\pi l_{\alpha\gamma}^2)$ and for numerical calculations it is specified in reciprocal k-space as

$$\omega_{\alpha\gamma}(k) = j_0\left(k l_{\alpha\gamma}\right), \tag{2.6}$$

where $j_0(x)$ is the zeroth-order spherical Bessel function. The intermolecular term in (2.5) to be input into (2.1) is given by the site-site radial total correlation function $h_{\alpha\gamma}(r)$ between all sites α and γ on all sorts of molecules in bulk solvent. The latter are obtained in advance to the 3D-RISM-KH calculation from the dielectrically consistent RISM theory [65, 66] coupled with the KH closure (DRISM-KH approach) [12, 14] which can be applied to a bulk solution of a given composition, including polar solvent, co-solvent, electrolyte, and ligands at a given concentration. The DRISM integral equation reads [65, 66]

$$\tilde{h}_{\alpha\gamma}(r) = \tilde{\omega}_{\alpha\mu}(r) * c_{\mu\nu}(r) * \left[\tilde{\omega}_{\nu\gamma}(r) + \rho_\nu \tilde{h}_{\nu\gamma}(r)\right], \tag{2.7a}$$

where $c_{\alpha\gamma}(r)$ is the site-site direct correlation function of pure bulk solvent, and both the intramolecular correlation function $\tilde{\omega}_{\alpha\gamma}(r)$ and the total correlation function $\tilde{h}_{\alpha\gamma}(r)$ are renormalized due to a dielectric bridge correction in a particular analytical form [65, 66] that ensures the given phenomenological value as well as consistency of the dielectric constant determined through the three different routes related to the solvent-solvent, solvent-ion, and ion-ion effective interactions in electrolyte solution,

$$\tilde{\omega}_{\alpha\gamma}(r) = \omega_{\alpha\gamma}(r) + \rho_\alpha \zeta_{\alpha\gamma}(r)\,, \tag{2.7b}$$

$$\tilde{h}_{\alpha\gamma}(r) = h_{\alpha\gamma}(r) - \zeta_{\alpha\gamma}(r)\,. \tag{2.7c}$$

The renormalized dielectric correction enforcing the given phenomenological value of the dielectric constant and the proper orientational behavior and consistency of the dielectric response in the electrolyte solution is obtained in the analytical form specified in the reciprocal k-space as follows [65, 66],

$$\zeta_{\alpha\gamma}(k) = j_0(kx_\alpha)\, j_0(ky_\alpha)\, j_1(kz_\alpha) h_{\mathrm{c}}(k)\, j_0(kx_\gamma)\, j_0(ky_\gamma)\, j_1(kz_\gamma)\,, \tag{2.8}$$

where $j_0(x)$ and $j_1(x)$ are the zeroth- and first-order spherical Bessel functions, $\mathbf{r}_\alpha = (x_\alpha, y_\alpha, z_\alpha)$ are the Cartesian coordinates of partial charge q_α of site α on species s with respect to its molecular origin; both sites α and γ are on the same species s, and its dipole moment $\mathbf{d}_s = \sum_{\alpha\in s} q_\alpha \mathbf{r}_\alpha$ is oriented along the z-axis: $\mathbf{d}_s = (0, 0, d_s)$. Note that the renormalized dielectric correction (2.8) is nonzero only for polar solvent species of electrolyte solution which possess a dipole moment and thus are responsible for the dielectric response in the DRISM approach. The envelope function $h_{\mathrm{c}}(k)$ has the value at $k = 0$ determining the dielectric constant of the solution and is assumed in a smooth non-oscillatory form quickly falling off at wavevectors k larger than the inverse characteristic size l of liquid molecules [65, 66],

$$h_{\mathrm{c}}(k) = A \exp\left(-l^2 k^2/4\right), \tag{2.9}$$

where $A = (y^{-1}\epsilon - 3)\rho_{\mathrm{polar}}^{-1}$ is the amplitude related to the dielectric constant ϵ of the electrolyte solution which is a phenomenological parameter specified at input. The form (2.8) has also been extended to solvent mixtures [67], where the total number density of polar species $\rho_{\mathrm{polar}} = \sum_{s\in\mathrm{polar}} \rho_s$ and the dielectric

susceptibility of the mixture is given in the general case by

$$y = \frac{4\pi}{9k_{\mathrm{B}}T} \sum_{a \in \mathrm{polar}} \rho_a d_a^2 . \tag{2.10}$$

The parameter l in the envelope function (2.9) specifies the characteristic separation from a molecule in solution below which the dielectric correction (2.8) is switched off so as not to distort the short-range solvation structure. It can be chosen to about $l = 1$ Å for water solvent; however, in solvent of larger molecules or in the presence of such co-solvent, it should be increased so as to avoid "ghost" associative peaks appearing in the radial distributions if the dielectric correction (2.8) interferes with the intramolecular structure of the large solvent species. For example, for octanol solvent, it should be set to about $l = 10$ Å.

The DRISM integral equation (2.7a) is complemented with the KH closure [12, 14] which in the site-site radial version reads

$$g_{\alpha\gamma}(r) = \begin{cases} \exp\left[-u_{\alpha\gamma}(r)/(k_{\mathrm{B}}T) + h_{\alpha\gamma}(r) - c_{\alpha\gamma}(r)\right] & \text{for} \quad g_{\alpha\gamma}(r) \leqslant 1, \\ 1 - u_{\alpha\gamma}(r)/(k_{\mathrm{B}}T) + h_{\alpha\gamma}(r) - c_{\alpha\gamma}(r) & \text{for} \quad g_{\alpha\gamma}(r) > 1, \end{cases} \tag{2.11}$$

where the site-site interaction potential $u_{\alpha\gamma}(r)$ (all-atom force field) is typically modelled with Coulomb and Lennard-Jones terms. The DRISM-KH equations (2.7a) and (2.11) keep the same dielectrically consistent asymptotics (2.8) as the original DRISM-HNC theory [65, 66], but extend the description to solutions with strong associative species in a wide range of compositions and thermodynamic conditions not amenable to the HNC closure. (The latter overestimates the associative attraction and density inhomogeneities and usually diverges for such systems.)

2.4. Analytical expressions for solvation thermodynamics

Much as the HNC approximation, the 3D-KH closure (2.4) to the 3D-RISM integral equation (2.1) has an exact differential of the solvation free energy, and thus allows one to analytically perform Kirkwood's thermodynamic integration gradually switching on the solute-solvent interaction. This gives the solvation free energy of the solute molecule (or supramolecule) in multicomponent solvent in a closed analytical form in terms of the 3D site total and direct correlation functions $h_\gamma(\mathbf{r})$ and $c_\gamma(\mathbf{r})$ [9, 12, 14]:

$$\Delta\mu = \sum_\gamma \int_V \mathrm{d}\mathbf{r}\, \Phi_\gamma^{\mathrm{KH}}(\mathbf{r}) , \tag{2.12a}$$

$$\Phi_\gamma^{\mathrm{KH}}(\mathbf{r}) = \rho_\gamma k_{\mathrm{B}}T \left[\frac{1}{2}\big(h_\gamma(\mathbf{r})\big)^2 \Theta\big(-h_\gamma(\mathbf{r})\big) - \frac{1}{2} h_\gamma(\mathbf{r}) c_\gamma(\mathbf{r}) - c_\gamma(\mathbf{r}) \right] , \tag{2.12b}$$

where the sum goes over all the sites of all solvent species, and $\Theta(x)$ is the Heaviside step function. This offers several substantial advantages over molecular simulations using free energy perturbation techniques with multiple productive runs to perform the thermodynamic integration or solute mutation, which is extremely time consuming [68]. Once the 3D-RISM-KH equations (2.1), (2.4) are converged for the correlation functions $c_\alpha(\mathbf{r})$ and $h_\alpha(\mathbf{r})$, all the solvation thermodynamics is readily available from the expression (2.12) and analysis of its components. Note that accurate calculation of the solvation free energy from the functional (2.12) requires proper analytical account of electrostatic asymptotics in all the correlation functions, both 1D and 3D, in the integral equations (2.1)–(2.11) [10–14].

The integrand $\Phi_\gamma^{\mathrm{KH}}(\mathbf{r})$ in (2.12a) is interpreted as the solvation free energy density (3D-SFED) coming from interaction site γ of solvent molecules around the solute. The solvation free energy of the solute macromolecule $\Delta\mu$ is obtained by summation of the 3D-SFED partial contributions from all solvent species and spatial integration over the whole space. Note that the integration volume V in the form (2.12) comprises both the solvation shells and the solute-solvent molecular repulsive cores, since the direct correlation functions $c_\alpha(\mathbf{r})$ inside the latter are related to the free energy of creation of a cavity to accommodate the solute excluded volume. In this way, $\Phi_\gamma^{\mathrm{KH}}(\mathbf{r})$ characterizes the intensity of effective solvation forces in different 3D spatial regions of the solvation shells and indicates where they contribute the most/least to the entire solvation free energy.

The solvation free energy in the form (2.12) can be split up into components from each solvent species by grouping the terms for the corresponding sites. Further, integrating the 3D-SFED values $\Phi_\gamma^{\mathrm{KH}}(\mathbf{r})$ over space regions geometrically attributed to the constituting chemical groups of the solute by a Voronoi decomposition (including the repulsive core region) resolves partial contributions of functional groups of the solute molecule (or supramolecule) to the solvation free energy, which is called spatial decomposition analysis [69, 70].

With the solvation free energy obtained in the form (2.12), the potential of mean force $w(\mathbf{1},\mathbf{n})$ between $i = 1, n$ solute molecules (or equally, supramolecule parts or nanoparticles) in solution at given distances and orientations in an arrangement $(\mathbf{1},\mathbf{n})$ is calculated directly by definition as their interaction potential $u(\mathbf{1},\mathbf{n})$ plus the change in the solvation free energy from the sum of $\Delta\mu(\mathbf{i})$ of separate particles to non-additive $\Delta\mu(\mathbf{1},\mathbf{n})$ of the particles brought together [10–14],

$$w(\mathbf{1},\mathbf{n}) = u(\mathbf{1},\mathbf{n}) + \Delta\mu(\mathbf{1},\mathbf{n}) - \sum_{i=1} \Delta\mu(\mathbf{i})\,. \tag{2.13}$$

Similarly to the solvation free energy (2.12), the PMF (2.13) can be split up by spatial decomposition analysis into contributions of solute chemical functionalities and into terms mediated by each solvent species. This route provides an efficient and accurate statistically representative way to calculate effective interactions of various chemical species [10–14, 27, 28], nanomaterials [30–35], and biomolecular nanosystems[53–57].

Furthermore, PMFs between a solute molecule / supramolecule / nanoparticle and solvent molecules averaged over molecular orientations centered at solvent interaction sites can be conveniently obtained in the 3D site representation from the definition of the 3D site density distribution functions,

$$w_\gamma(\mathbf{r}) = -k_\mathrm{B}T \ln\left[g_\gamma(\mathbf{r})\right]. \tag{2.14}$$

The solute-solvent PMFs so obtained in a single 3D-RISM-KH calculation provide 3D spatial maps of site binding affinity of solvent species to the solute molecule (positive in attractive wells and negative in repulsive barriers of PMFs). This constitutes a new approach toward molecular recognition and computational fragment-based drug design [41]. With fragmental decomposition of flexible ligands treated as distinct species in solvent mixture of arbitrary complexity, the computed density functions and the corresponding PMFs (2.14) of solvent, ions, and ligand fragments around a biomolecule obtained from the 3D-RISM-KH theory on discrete 3D spatial grids uniquely define the scoring function which can be interfaced to a docking protocol [41, 42, 57, 58], e.g., in the 3D-RISM-Dock approach implemented in the AutoDock suite for automated ranking of docked conformations [58]. As a validation, the 3D-RISM-Dock tool predicted the location and residency times of the modes of binding of a flexible thiamine molecule to the prion protein at near-physiological conditions in an excellent agreement with experiment [57, 58]. The capabilities of 3D maps of ligand binding affinity calculated from (2.14) include the prediction of functions in biomolecular systems, such as drug efflux pathway in multidrug transporter AcrB [59].

Further, the solvation free energy (2.12) can be split up into the energetic and entropic contributions [3, 71, 72],

$$\Delta\mu = \Delta\varepsilon^{\mathrm{uv}} + \Delta\varepsilon^{\mathrm{vv}} - T\Delta s_V\,, \tag{2.15}$$

by calculating the solvation entropy at constant volume as

$$\Delta s_V = -\frac{1}{T}\left(\frac{\partial\Delta\mu}{\partial T}\right)_V \tag{2.16}$$

and the internal energy of the solute (“u”) — solvent (“v”) interaction as

$$\Delta\varepsilon^{\mathrm{uv}} = -k_\mathrm{B}T\sum_\gamma \int \mathrm{d}\mathbf{r}\, g_\gamma(\mathbf{r})\, u_\gamma(\mathbf{r})\,, \tag{2.17}$$

where the remaining term $\Delta\varepsilon^{\mathrm{vv}}$ gives the energy of solvent reorganization around the solute molecule. In the same way, the PMF (2.13) or (2.14) can be decomposed into energetic and entropic terms.

Further to the solvation free energy, the 3D-RISM-KH theory also provides analytical expressions for other thermodynamic functions, including the partial molar volume (PMV) of the solute molecule obtained in the Kirkwood-Buff theory in terms of the 3D direct correlation functions from the 3D-RISM integral equation [48–52],

$$\bar{V} = k_{\mathrm{B}} T \chi_T \left(1 - \sum_{\gamma} \rho_{\gamma} \int \mathrm{d}\mathbf{r}\, c_{\gamma}(\mathbf{r}) \right), \tag{2.18}$$

where χ_T is the isothermal compressibility of bulk solvent which is analytically obtained in the so-called compressibility route from the RISM integral equation as [23]

$$k_{\mathrm{B}} T \chi_T = \rho^{-1} \left(1 - 4\pi \sum_{\alpha\gamma} \rho_{\alpha} \int_0^{\infty} r^2 \mathrm{d}r\, c_{\alpha\gamma}(r) \right)^{-1} \tag{2.19}$$

in terms of the site-site direct correlation functions of bulk solvent $c_{\alpha\gamma}(r)$ calculated along with the solvent susceptibility $\chi_{\alpha\gamma}(r)$ from the DRISM-KH equations, and $\rho = \sum_s \rho_s$ is the total number density of species s in the solution bulk.

As noted above, the solvation free energy expression (2.12) follows from the functional form of the KH closure (2.4) to the 3D-RISM integral equation (2.1) by carrying out the thermodynamic integration analytically. In practice, the 3D-RISM-KH theory can be used with a different functional of the solvation free energy. In particular, Chandler and co-workers derived the so-called Gaussian Fluctuation (GF) free energy functional based on the assumption of Gaussian fluctuations for the solvent, which is similar to the HNC functional form but without the h_{γ}^2 term [73, 74]. The solvation free energy calculated from the GF functional using site-site correlation functions obtained from the RISM-HNC integral equations were generally in reasonable agreement with experiment and in better agreement than those obtained with the HNC functional [75, 76]. In the context of 3D-RISM theory, the GF functional reads [36, 56]

$$\Delta\mu = \sum_{\gamma} \int_V \mathrm{d}\mathbf{r}\, \Phi_{\gamma}^{\mathrm{GF}}(\mathbf{r}), \tag{2.20a}$$

$$\Phi_{\gamma}^{\mathrm{GF}}(\mathbf{r}) = \rho_{\gamma} k_{\mathrm{B}} T \left[-\frac{1}{2} h_{\gamma}(\mathbf{r}) c_{\gamma}(\mathbf{r}) - c_{\gamma}(\mathbf{r}) \right]. \tag{2.20b}$$

The GF functional using the 3D correlation functions obtained from the 3D-RISM-KH theory provides the hydration free energy in better agreement with experiment in some cases [56, 77].

Recently, it was found that the 3D-RISM-KH theory supplemented with the solvation free energy correction based on the partial molar volume of the compound (also obtained from the 3D-RISM-KH theory) provides the hydration free energy for a large diverse set of compounds with an accuracy comparable to much more computationally demanding molecular dynamics simulations with explicit solvent [77]. The correction to the solvation free energy $\Delta\mu$ calculated from either the KH functional (2.12) or the GF one (2.20) is constructed as a linear function of the PMV in turn calculated in the form (2.18) using the correlation functions from the 3D-RISM-KH theory,

$$\Delta\mu_{\mathrm{UC}} = \Delta\mu + \alpha\left(\rho\bar{V}\right) + \beta. \tag{2.21}$$

The constants α and β in (2.21) are obtained from the linear regression analysis. This partial molar volume correction technique, also called a “Universal Correction” (UC), was parameterized with the GF functional on hydration free energies of a training set of 65 molecules and was tested on a set of 120 molecules [77]. The UC to the KH functional has also been parameterized for the hydration free energy on a set of 504 organic molecules [78]. In the same study, another correction to the KH functional based on the Ng modified free energy functional [79] was introduced and shown to accurately predict hydration free energies of the same set of 504 compounds [78].

With the partial molar volume correction (or the UC term) re-parameterized accordingly, the 3D-RISM-KH theory also provides an excellent agreement with the experimental data for the solvation free energy in non-polar solvent (1-octanol) of a large library of 172 small compounds with diverse functional groups, and so accurately predicts the octanol-water partition coefficients [25]. The best agreement with the experimental data for octanol-water partition coefficients is obtained with the KH-UC solvation free energy functional.

2.5. Analytical treatment of electrostatic asymptotics

To properly treat electrostatic forces in electrolyte solution with polar solvent and ionic species in 3D-RISM / DRISM theory requires analytical treatment of the electrostatic asymptotics of the radial site-site total and direct correlation functions of bulk solvent in the DRISM-KH equations (2.7a), (2.11), as well as of the 3D site total and direct correlation functions in the 3D-RISM-KH equations (2.1), (2.4), the solvation free energy functional (2.12) or (2.20), its derivatives (2.13)–(2.17), and the volumetrics (2.18)–(2.19) [10–14, 18]. The convolution in the 3D-RISM integral equation (2.1) is calculated by using the 3D fast Fourier transform (3D-FFT) technique. The DRISM-KH equations are discretized on a uniform radial grid of resolution $0.01 - 0.1$ Å, and the 3D-RISM-KH equations are discretized on a uniform 3D rectangular grid with resolution $0.2 - 0.5$ Å in a 3D box of size including 2–3 solvation shells around the solute (supra)molecule. (Typical box grid sizes range from $64\times64\times64$ to $256\times256\times256$ nodes.) Analytical forms of the non-periodic electrostatic asymptotics in the direct and reciprocal space are separated out from all the correlation functions before and then added back after the 3D-FFT. Note that although the solvent susceptibility $\chi_{\alpha\gamma}(r)$ has a long-range electrostatic part, no aliasing occurs in the backward 3D-FFT of the short-range part of $h_\gamma(\mathbf{k})$ on the 3D box supercell since the short-range part of the convolution product $h_\gamma(\mathbf{r})$ contains merely 2–3 oscillations (solvation shells) and vanishes outside the 3D box for physical reasons [13, 14, 18]. Accordingly, the electrostatic asymptotics terms in the thermodynamic integral (2.12) are handled analytically and reduced to one-dimensional integrals easy to compute [10–14, 18].

2.6. Accelerated numerical solution

The 3D-RISM-KH integral equations (2.1), (2.4) are converged typically to a root mean square accuracy of $10^{-4} - 10^{-8}$ and the DRISM-KH equations (2.7a), (2.11) to an accuracy of $10^{-6} - 10^{-12}$ by using the modified direct inversion in the iterative subspace (MDIIS) accelerated numerical solver [10–13, 80]. MDIIS is an iterative procedure accelerating the convergence of integral equations of liquid state theory by optimizing each iterative guess in a Krylov subspace of typically last 10–20 successive iterations and then making the next iterative guess by mixing the optimized solution with the approximated optimized residual [10, 12, 13, 80]. It is closely related to Pulay's DIIS approach for quantum chemistry equations [81]. and other similar algorithms like the generalized minimal residual (GMRes) solver [82]. The MDIIS solver combines the simplicity and relatively small memory usage of an iterative approach with the efficiency of a direct method. Compared to damped (Picard) iterations, MDIIS provides substantial acceleration with quasiquadratic convergence practically throughout the whole range of root mean square residual, and is robust and stable. Of particular importance is that MDIIS ensures convergence (provided a solution exists) for complex charged systems with strong associative and steric effects, which is usually not achievable with Picard iterations and constitutes a challenging task in the case of 3D integral equations on large 3D grids. Converging the 3D-RISM-KH equations takes minutes to hours on a HPC workstation, depending on the 3D grid size, whereas the DRISM-KH equations converge in seconds for water but much slower, up to hours, for multi-component solvent mixtures with complex species.

3. Examples of 3D-RISM-KH calculations for the solvation structure and thermodynamics of solution systems and nanoparticles

Sketched below are the capabilities of the 3D-RISM-KH molecular theory of solvation in predictive modelling of some representative molecules and nanosystems in solution: structure and potentials of mean force in aqueous electrolyte solutions, activity coefficients of simple and molecular ions in electrolyte solutions, structural transitions in mixtures of water and amphiphile co-solvent, compound partitioning between water and octanol solvents, and effective nanoscale forces between cellulose nanoparticles in hemicellulose hydrogel maintaining the resilient structure of plant cell walls. The discussion below concerns the main features, capabilities and conclusions, whereas the force fields used and different setup details can be found in the corresponding original work cited here.

3.1. Simple ions in aqueous electrolyte solution

One of the basic cases of electrolyte solution systems is aqueous solution of sodium chloride. Figure 1 presents the 3D-RISM results for the solvation structure and potential of mean force for all the pairs of Na^+ and Cl^- ions in aqueous electrolyte solution at infinite dilution and at a relatively high concentration of 1.069 mol/l [10, 11]. This generic system presents an illustrative test for a solvation theory to reproduce most of the essential features of a variety of chemical and biomolecular effects in solution, and the 3D-RISM theory succeeds in that.

The 3D site distributions of the water oxygen (O) and hydrogen (H) site distributions around the pair of the Na^+ and Cl^- ions in aqueous solution at infinite dilution are shown in figure 1 (a). Both the O and H distributions form high crowns of the first solvation shell around the ions, with the high O and lower H peaks near the contacts of the crowns corresponding to water molecules bridging the ions. This structure is then followed by the shallow second solvation shells. The corresponding potential of mean force (PMF) calculated using equation (2.13) is shown in figure 1 (c). The H distribution has two crowns around the Cl^- ion; their separation from the ion shows that the inner H crown gives the water hydrogens in contact with the Cl^- ion while the outer H crown corresponds to the other water hydrogens looking outwards but at the angle determined by the tetrahedral hydrogen bonding structure of the water. On the other hand, there is a single H crown around the Na^+ ion and the separations of the O and H crowns from Na^+ show that both the water hydrogens are looking outward Na^+, tilted at the same angle. These are typical arrangements of water around a cation oriented like a dipole and water bonded with one hydrogen to an anion. The arrangements are visualized in the cartoons in figure 1 (b) for the 3D water site distributions around both the contact ion pair (CIP) and solvent separated ion pair (SSIP) of the Na^+ and Cl^- ions. Water molecules in contact with both the cation and anion form the dipole like association with the former and the hydrogen bonding with the latter, thus creating a water bridge of strongly associated water molecules located in a ring between the ions. This bridge strongly deepens the solvation contribution to the PMF at the CIP arrangement; interplaying with the ion-ion core repulsion, this results in a significant shift of the first minimum on the PMF to a shorter ion-ion separation, compared to the primitive solvation model just uniformly reducing the Coulomb attraction between the ions by the water dielectric constant $\epsilon = 80$ [dashed line in figure 1 (c)]. At the SSIP arrangement, the water bridge strengthens the association between the ions and results in the second minimum on the ion-ion PMF [figure 1 (d)]. Oscillations diminish with distance and the PMF goes to the limit of the pure dielectrically screened electrostatic potential (dashed line). At an intermediate separation between the ions, a desolvated gap forms due to the steric effect of expulsion of solvent molecules by the repulsive cores of the ions [figure 1 (b)]. The work against the solvent environment to expel the solvent and create the desolvation gap results in a barrier between the PMF first and second minima corresponding to the CIP and SSIP arrangements [figure 1 (c)].

The decomposition of the PMFs for the pairs of Na^+ and Cl^- ions is also shown in figure 1 (e), (f). The non-electrostatic (or non-polar) part of the PMFs can be defined either as a difference of the full PMF and its Born electrostatic part (solid curve without symbols in the middle right-hand part for the Na^+–Cl^- ion pair) or as a PMF obtained from the 3D-RISM calculations for the ion pair solute with the same Lennard-Jones short-range potentials but with the charges of the ions entirely switched off while keeping the charges of water molecules (solid curve with symbols in the middle right-hand part). The difference between these two curves illustrates a strong coupling and interplay of the electrostatic and non-electrostatic forces in the solvation structure. Note that in an aqueous solution or other highly polar solvent, the electrostatic part of the mean solvation forces responsible for dielectric screening always opposes and almost cancels out the Coulomb interaction in the ionic pair and so the resulting PMF comes out effectively scaled down by the high dielectric constant of the polar solvent.

The PMFs in the Na^+–Na^+ and Cl^-–Cl^- pairs of like ions have the same features of the first and second minima and the barrier between them due to the interplay of the associative forces and molecular structure of the ions and solvent molecules. However, at infinite dilution, the strength of the solvent bridges is not sufficient to overcome the electrostatic repulsion and to stabilize the like ion pairs, and both the first and second minima are local [figure 1 (c)]. (Note that this can be very different for large molecular ions with weaker electrostatic attraction at the separation determined by their size.) The picture changes for the electrolyte solution at a high concentration of 1 mol/l; numerous salt bridges form in addition to water hydrogen bridges, and the like ion pairs get stabilized in both CIP and SSIP arrangements [figure 1 (d)].

For example, the Cl^-–Cl^- ion pair in the aqueous solution at this concentration is bridged by several Na^+ ions and water molecules, forming a cluster depicted in figure 1 (g). The structure of such a cluster follows from the analysis of the 3D-RISM results for the 3D distribution functions of solution species and the corresponding coordination numbers of Na^+, Cl^-, and water around each ion pair (Na^+–Cl^-, Cl^-–Cl^-, and Cl^-–Cl^-).

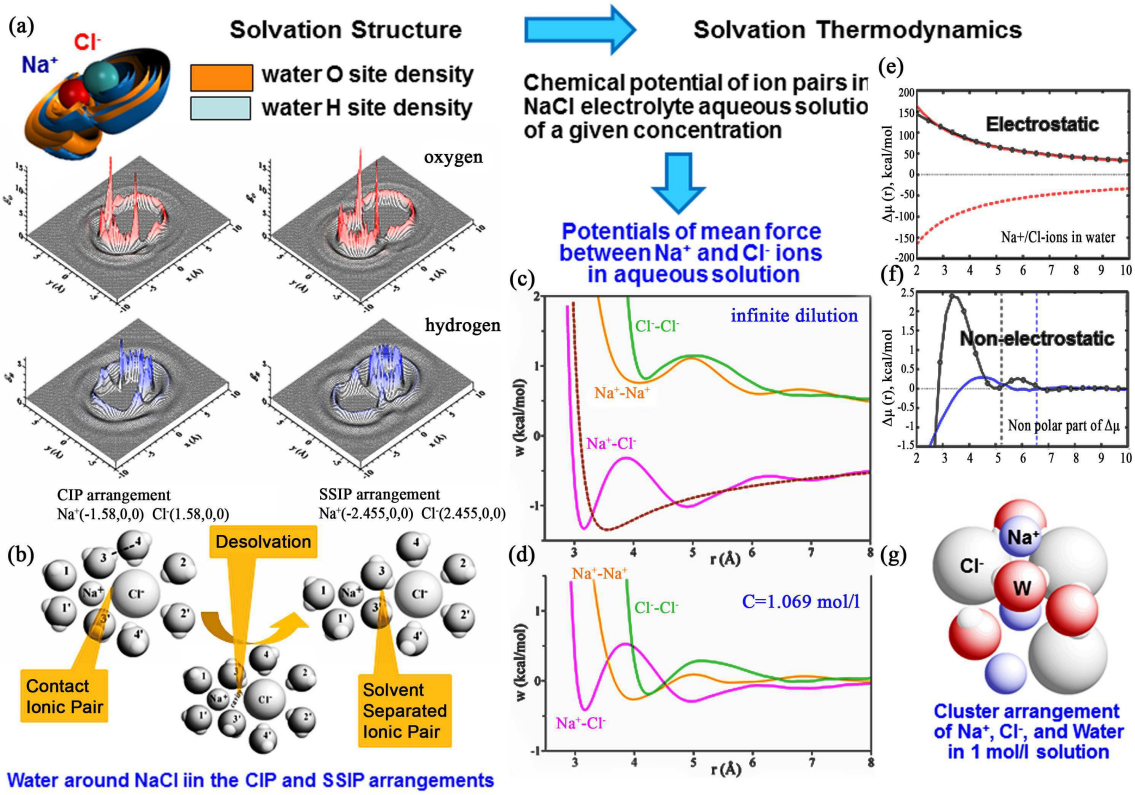


Figure 1. (Color online) Solvation structure of the contact ion pair (CIP) and solvent separated ion pair (SSIP) of the Na^+ and Cl^- ion pair in aqueous solution at infinite dilution, obtained from 3D-RISM theory [10, 11]. Part (a): Section of the 3D distribution functions of water oxygen (O) and hydrogen (H) in the plane passing through the ion-ion axis. Part (b): Visualization of the water solvent arrangements around the CIP and SSIP, as well as the ions separated by a gap corresponding to the barrier at their potential of mean force. Part (c): Potentials of mean force calculated using (2.13) for all the pairs of Na^+ and Cl^- ions in aqueous electrolyte solution at infinite dilution and at concentration 1.06 mol/l. Part (d): Shown for comparison is also the potential of mean force of the Na^+–Cl^- ion pair in the structureless dielectric continuum (primitive model) of water solvent. Decomposition of the potential of mean force of the ion pairs at infinite dilution into: (d) electrostatic term for the Na^+–Na^+, Cl^-–Cl^-, and Na^+–Cl^- ion pairs (two upper solid curves and lower dashed curve, respectively); (e) non-electrostatic (non-polar) term for the Na^+–Cl^- ion pair at full ion charges and with ion charges switched off (solid curves with and without symbols, respectively). Part (g): Visualization of a cluster hydrated ions, based on the peaks positions on the 3D maps of density distributions: Cl^- ions (shown in gray) in a CIP arrangement stabilized due to bridging by associated Na^+ ions (in blue) as well as by hydrogen bonded water molecules (water O in red, water H in white) in aqueous electrolyte solution of concentration 1M.

3.2. Activities of ions in electrolyte solution

The predictive capability of the 3D-RISM-KH theory (DRISM-KH for simple ions as solutes) in reproducing the solvation thermochemistry of electrolyte solutions in a wide range of concentrations has been validated and shown to be in good agreement with experiment, in particular, for aqueous electrolyte solutions of molecular as well as simple ions in ambient conditions at concentrations from infinite dilution

to high ionic strength [64, 83]. In particular, the theory yields the activity coefficient of the ion γ_i in terms of its solvation free energy $\Delta\mu(c)$ in solution with solvent density $\rho_s(c)$ at electrolyte concentration c with respect to those at infinite dilution $c = 0$ [64, 83],

$$\gamma_i = \frac{\rho_s(c)}{\rho_s(c=0)} \exp\left[\frac{\Delta\mu(c) - \Delta\mu(c=0)}{k_B T}\right]. \tag{3.1}$$

For illustration, figure 2 presents a comparison of the activity coefficient in ambient aqueous solution of sodium chloride at concentrations up to 1 mol/kg calculated from the DRISM-KH theory against experimental data. The calculated activity coefficient is in a good agreement with experiment in the whole range of concentrations.

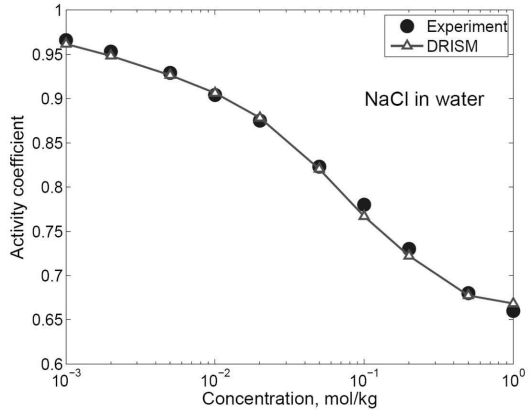


Figure 2. Activity coefficients of ions in ambient aqueous solution of sodium chloride at concentration from low to high against experimental data.

3.3. Structure of water-alcohol mixtures

A further important example of crucial importance in chemistry and biomolecular nanosystems is the formation of nanostructures in solution driven by hydrophobic attraction. Figure 3 illustrates the RISM theory predictions for the structure of the ambient mixtures of water and tert-butyl alcohol (TBA) in the whole range of concentrations [23, 24]. TBA is a generic example of primitive surfactant with a hydrophobic head of four carbons and a hydrophilic "tale" represented by the hydroxyl group. The solvation structure of this system successively goes through several stages with TBA concentration in water changing from infinite dilution pure TBA: a separate TBA molecule embedded in a water tetrahedral hydrogen bonding cage at infinite dilution changes to micromicelles of four to six TBA molecules in the head-to-head arrangement incorporated in a water hydrogen bonding cage at about 4% TBA molar fraction; then, the tetrahedral hydrogen bonding structure of water gets disrupted at about 40% TBA molar fraction to be replaced by the zigzag hydrogen bonding structure of alcohol, with separate water molecules embedded in it at infinite dilution of water in TBA. The RISM theory predicted both the structure and thermodynamics of these mixtures, in particular, the concentration and temperature dependence of the compressibility, including the isosbestic point and minimum corresponding to the formation of micromicelles [23, 24], in an excellent agreement with the findings from neutron diffraction experiment [84, 85] and compressibility measurements (see references in [24]).

3.4. Solvation free energy of small compounds in water and octanol

The octanol-water partition coefficient characterizes hydrophobic (lipophilic) and hydrophilic properties of chemical compounds [86, 87]. For dilute solutions it is defined as the ratio of molar concentrations of a compound in octanol and water, $P_{O/W} = [C_O]/[C_W]$. In practice, it is more common to use logarithm of the molar concentration ratio due to the large range of changes of the partition coefficient.

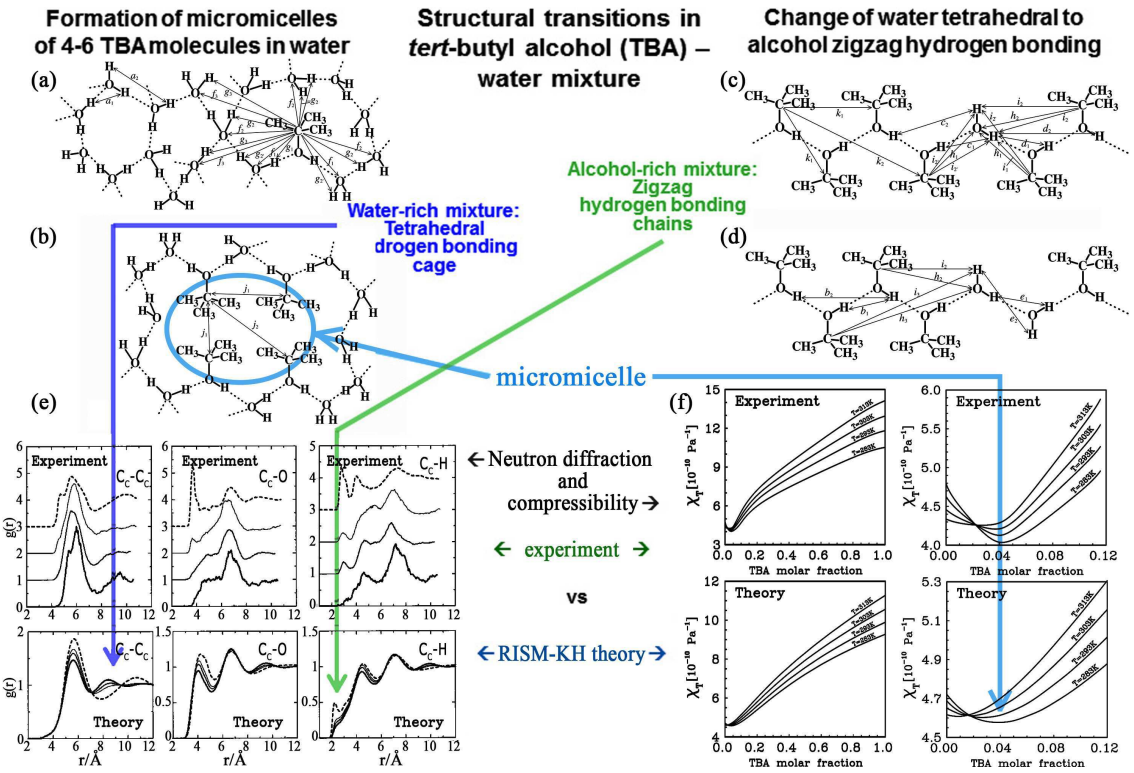


Figure 3. (Color online) Structural transitions, solvation structure and isothermal compressibility with concentration in a mixture of water and tert-butyl alcohol (TBA) predicted by the RISM-KH molecular theory of solvation [23, 24] versus experimental findings [84, 85] (and references in [24]). Part (a): tert-butyl alcohol molecule at infinite dilution incorporated in a cage of water tetrahedral hydrogen bonding; part (b): micromicelle of four tert-butyl alcohol molecules in a cage of the water hydrogen bonding. Part (c): water molecules incorporated in the alcohol zigzag hydrogen bonding structure at infinite dilution of water; part (d): the same as in part (c), but at low water concentration. Part (e): solvation structure features corresponding to the water tetrahedral hydrogen bonding at low alcohol – high water and the alcohol zigzag hydrogen bonding at high alcohol – low water concentrations. Part (f): isosbestic point and minimum of the compressibility corresponding to the formation of TBA micromicelles in water.

As an equilibrium constant, the partition coefficient is directly related to the transfer free energy of the compound from water (polar solvent) to octanol (non-polar solvent),

$$\log P_{\mathrm{O/W}} = \left(\Delta\mu_{\mathrm{W}} - \Delta\mu_{\mathrm{O}}\right) / \left(k_{\mathrm{B}}\, T\right) . \tag{3.2}$$

The partition coefficient is one of the most important physico-chemical characteristics used in pharmacology, environmental studies, food industry, etc. In particular, in pharmacology, the partition coefficient is used to predict distribution of drugs within the body. It is a crucial parameter that defines drug-likeness of a compound. The partition coefficient, along with other descriptors, defines the efficiency of a drug crossing the blood brain barrier and reaching the central nervous system [88].

Due to the practical importance, many theoretical methods have been developed to predict the partition coefficient, ranging from the phenomenological approaches based on continuum solvation models or different descriptors such as quantitative structure-activity relationship to the sophisticated approaches based on first principle calculations of the transfer free energy between different solvents, including explicit solvent MD simulations and calculations based on different solvation models with account for quantum mechanical effects (see references in [25]). A major drawback of the continuous solvation methods is their inability to treat specific solute-solvent and solvent-solvent interactions such as hydrogen bonding, and their non-transferability (re-parametrization is required for a new solvent composition and possibly thermodynamic conditions). In principle, these difficulties can be avoided by using all-atom explicit sol-

vent MD simulations. However, such simulations are computationally demanding and cannot be used in most practical applications such as virtual screening in rational drug design. An alternative to explicit solvent MD simulations, the 3D-RISM-KH molecular theory of solvation with the partial molar volume correction provides an excellent agreement with the experimental data for the solvation free energy in both water and 1-octanol, and so accurately predicts the octanol-water partition coefficient [25].

The 3D correlation functions of solvent around the solute compound are obtained by converging the 3D-RISM-KH integral equations (2.1), (2.4), and then used to calculate the solvation free energy from the KH functional (2.12) supplemented with the UC term (2.21) in turn calculated from the partial molar volume (2.18) using the 3D-RISM-KH correlation functions. Figures 4 and 5 illustrate the performance of the 3D-RISM-KH theory using the KH solvation free energy functional (2.12) without and with the UC term (2.21) on solvation free energies in water and octanol [25] for a large set of small compounds with diverse chemical groups against experimental data [89, 90]. The linear coefficients in the UC expression (19) parameterized for the solvation free energy in water and octanol calculated with the KH solvation free energy functional are displayed in Table 3.4. The use of the UC to the KH free energy functional significantly improves the accuracy of the solvation free energy obtained from the 3D-RISM-KH theory (root mean square error (RMSE) drops from 22.84 to 1.975 kcal/mol), and thus makes it comparable to the results of explicit solvent MD simulations [77]. A considerably better performance of the 3D-RISM-KH theory with the KH free energy functional alone is observed in the case of solvation in octanol compared to hydration. Even in this case, the use of the UC further improves the accuracy of the results compared to experimental data (the RMSE from 1.37 down to 1.37 kcal/mol, see Table 3.4).

Table 1. Linear coefficients in the universal correction (UC) (2.21) to the Kovalenko-Hirata solvation free energy functional (2.12) for the two solvents [25]. Shown is the root mean square error (RMSE) between the calculated results (with and without UC) and experimental data for the library compounds [89, 90].

Solvent	α (kcal/mol)	β (kcal/mol)	RMSE (kcal/mol) with UC (without UC)
Water	–4.58	0.340	1.975 (22.84)
Octanol	0.083	–0.939	1.03 (1.37)

Finally, figure 6 makes a comparison of the octanol-water partition coefficient predicted from the 3D-RISM-KH theory with the KH free energy functional supplemented with the UC term for the library of compounds76 against experimental data [89, 90]. Shown for comparison are the results obtained for this library of small compounds by using the generalized Born solvent accessible surface area (GBSA) continuum solvation model widely popular in biomolecular calculations [25]. Compared to experimental data, the prediction accuracy level of 3D-RISM-KH with UC reaches the root mean square error as low as 1.28 which is almost twice better than RMSE of 2.32 for the GBSA results.

3.5. Effective interactions between cellulose nanoparticles in hydrogel

One representative illustration of the predictive capabilities of the 3D-RISM-KH molecular theory of solvation comes from the quest for fundamental understanding of the chemically-driven effective nanoscale forces that maintain plant cell wall structure [33–35]. Efficient conversion of lignocellulosic biomass to second-generation biofuels and valuable chemicals requires decomposition of resilient cell wall structure of plants. Overcoming biomass recalcitrance constitutes the most fundamental unsolved problem of plant-based green technologies [91, 92]. Plants naturally evolved to withstand harsh external mechanical, thermal, chemical, and biological factors. Their secondary cell walls are composed of cellulose microfibrils embedded in a complex non-cellulosic matrix comprised mainly of hemicellulose and lignin which are responsible for the cohesive forces within the cell wall that entail structural support to plants [93]. Current technological applications demand decomposition of this resilient structure in order to extract cell wall components for production of second-generation biofuels and other valuable chemical commodities. It is well known that cell wall recalcitrance varies among plant species and even within different phenotypes of the same plant. Close relationship between recalcitrance and chemical composition of a non-cellulosic matrix suggests that cell wall strength could be tuned by carefully controlling the

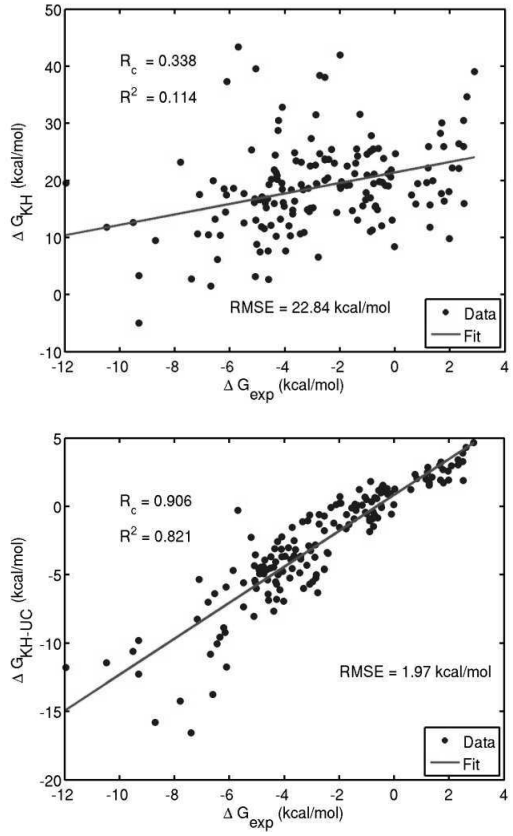


Figure 4. Hydration free energies obtained from the KH solvation free energy functional without (top) and with (bottom) the UC term using the correlation functions from the 3D-RISM-KH theory [25]. A comparison against experimental data for the library compounds [89, 90].

matrix composition [91, 93–95]. Full understanding of the chemical interactions within the cell walls is thus fundamental to gain control of the lignocellulosic biomass recalcitrance [92].

MD simulations have been used to investigate decrystallization of cellulose [96, 97], the interactions between cellulose and non-cellulosic components of plant cell walls [98], and the structure and dynamics of lignin [99]. However, extremely time-consuming and costly simulations are required to obtain adequate statistical sampling, addressing both solvation structure and thermodynamics of effective interactions in cell walls based on molecular forces. The 3D-RISM-KH molecular theory of solvation bridges the gap between molecular structure and effective forces on multiple length scales, and is uniquely capable of predicting the chemistry-driven effective interactions in plant cell walls [33–35].

Glucuronoarabinoxylan — hemicellulose type most abundant in important lignocellulosic grasses used for biofuel production, such as sugar cane and corn — consists of a xylan backbone decorated with branches of mainly arabinose and glucuronic acid whose amount and ratio substantially varies with the plant genotype [100]. Genetic manipulation of glucuronic acid branching has been shown to significantly improve xylan extractability from cell walls without impairing plant growth [93]. All the chemical changes in the xylan structure regard the branches, mainly arabinose and glucuronic acid (figure 7, top).

The modelling of structure and stability of plant cell walls consisted of the following stages [33]. (i) Calculation began with constructing a model of a primary cell wall represented with two 4-chain 8-glucose-long cellulose fragments, or cellulose nanocrystallites (CNs), immersed in aqueous solution of arabinose, glucuronic acid, and glucuronate monomers at different concentrations (figure 7, bottom). The Iβ CNs were built by using the Cellulose-Builder[101] in such a way as to have both the hydrophobic and hydrophilic faces exposed to the solvent. (ii) With an input of the interaction potentials between the species in this cell wall model, the 3D-RISM-KH integral equations were solved for the 3D site correlation func-

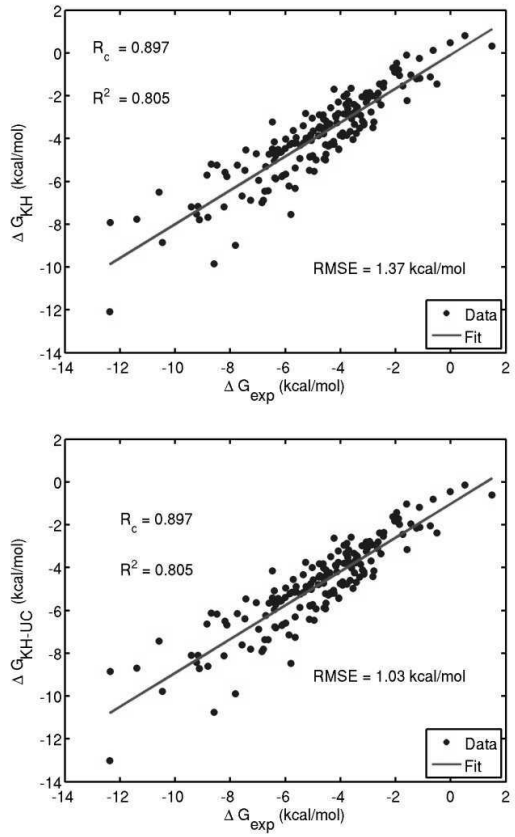


Figure 5. Solvation free energy in octanol obtained from the KH solvation free energy functional without (top) and with (bottom) the UC term using the correlation functions from the 3D-RISM-KH theory [25]. A comparison against experimental data for the library compounds [89, 90].

tions of hemicellulose monomers and water solvent ("hydrogel") in the solvation structure of two CNs at different separations and arrangements. (iii) The ensemble-averaged 3D spatial maps of the solvation free energy density (SFED) of the hemicellulose moieties around the CNs as well as the solvation free energies were obtained for each arrangement of the CNs. (iv) Finally, the PMFs between the CNs immersed in the hemicellulose hydrogel were calculated from (2.13).

Figures 8 (a) and 8 (b) depict the pathways of aggregation of CNs approaching each other with their hydrophilic faces (top row) and with their hydrophobic ones (bottom row), respectively. The corresponding PMFs against the separation between the CNs along these pathways are shown in figures 8 (c) and 8 (d), respectively. The PMFs for disaggregation of the CNs in both the hydrophilic (part a) and hydrophobic (part b) face arrangements exhibit two well defined local minima. In both cases, the first minimum at the separation marked as $r_{\mathrm{fc}} \approx 0$ Å corresponds to two aggregated CNs in direct face contact, whereas the PMF second minimum for CNs 3 Å apart refers to CNs separated by a solvent layer and reaches the well depth of -7 kcal/mol in pure water. These results support the suggestion that CNs aggregation through hydrophilic faces is preferable to the hydrophobic faces in primary cell walls [94].

The stability of aggregated CNs indicates how difficult it is to disrupt their arrangement within a primary cell wall. With increase of the glucuronate concentration, the face contact aggregation free energy $\Delta G_{\mathrm{agg}} = \mathrm{PMF}(r_{\mathrm{fc}})$ decreases similarly to the first maximum $\mathrm{PMF}(r_{\mathrm{bar}})$, while the disaggregation barrier $\Delta G_{\mathrm{dis}} = \mathrm{PMF}(r_{\mathrm{bar}}) - \mathrm{PMF}(r_{\mathrm{fc}})$ remains the same [figures 8 (c) and 8 (d)]. (The PMFs of CNs in glucuronic acid and arabinose hydrogel have similar shapes.) The larger the amount of hemicellulose, the stronger the primary plant cell wall microstructure. This is consistent with the fact that drastic structure change or absence of hemicellulose prevents plant growth due to structural collapse [100].

Figures 8 (e) and 8 (f) show the aggregation free energies (ΔG_{agg}) between two CNs along the hy-

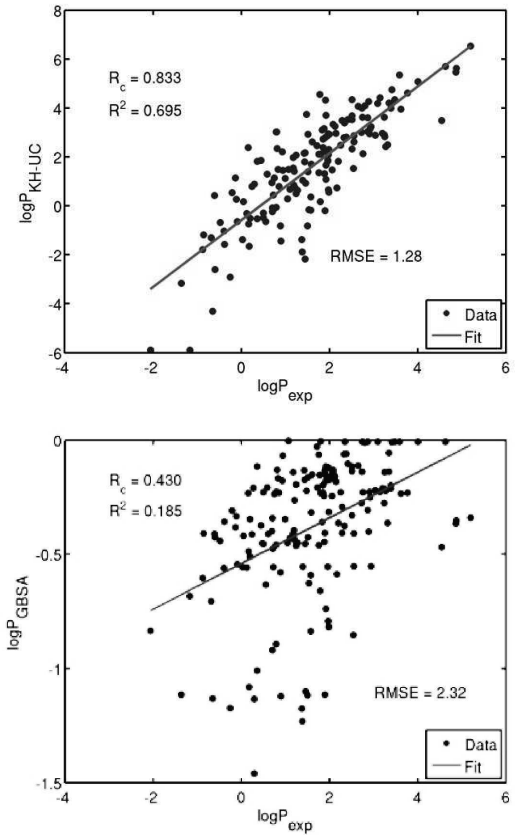


Figure 6. Logarithm of the octanol-water partition coefficient obtained from the KH solvation free energy functional with the UC term using the correlation functions from the 3D-RISM-KH theory (top) and from the GBSA continuum solvation model (bottom) [25]. A comparison against experimental data for the library compounds [89, 90].

drophilic (top) and hydrophobic face contacts (bottom). The hydrophilic face contact arrangement of the cellulose aggregate [figure 8(a)] is strongly stabilized due to interfibrillar hydrogen bonds and gives a global minimum of PMF. The global minimum for dissociation along the hydrophobic contact surface [figure 8 (b)] corresponds to the solvent-separated arrangement with the CNs ≈ 3 Å apart and the well depth reaching ≈ -7 kcal/mol in pure water.

Figure 9 makes a comparison of the PMFs for all types of hemicellulose monomers in aqueous solution: arabinose, glucuronic acid, and glucuronate. The changes observed are quantitative. The overall features of the PMF are preserved, whereas the magnitude of the changes is strongly affected by the concentration and nature of the hemicellulose solutions.

Finally, figure 10 shows the isosurfaces of the 3D spatial maps of the solvation free energy density (3D-SFED) coming from glucuronate at the cellulose surface. The solvation free energy (2.12a) of the cellulose nanofibril immersed in hemicellulose hydrogel is obtained by integrating the 3D-SFED contributions (2.12b) from all hydrogel species over the solvation shells. The glucuronate contribution in 3D-SFED varies from large negative values for the thermodynamically most favorable arrangements of glucuronate at the cellulose surface to small negative values for less favorable arrangements. The larger negative value isosurface [figure 10 (a)] is highly localized around polar sites on the cellulose surface and indicates hydrogen bonding of hemicellulose to cellulose. The smaller negative value isosurface [figure 10 (b)] indicates a diffuse second layer of hemicellulose monomers stacking over the cellulose surface. The stacking interactions in the second layer are weaker and less specific than hydrogen bonding and are due to hydrophobic as well as enhanced inter-molecular C–H…O interactions found in the crystal structure of cellulose [100]. Although the stacking interactions play a considerable role, the hemicellulose-

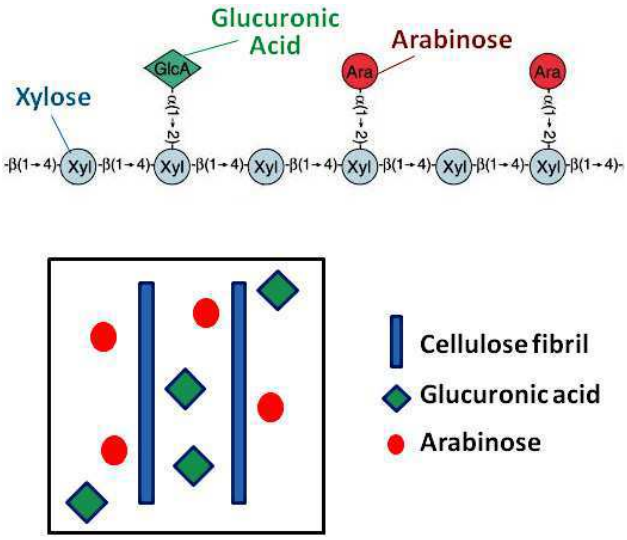


Figure 7. (Color online) Chemical structure of glucuronoarabinoxylan type hemicellulose: a xylan backbone decorated with branches of mainly arabinose and glucuronic acid in varying amount and ratio (top part). Model of plant cell walls represented with cellulose nanofibrils immersed in aqueous solution of arabinose, glucuronic acid, and glucuronate monomers at different concentrations (hemicellulose "hydrogel").

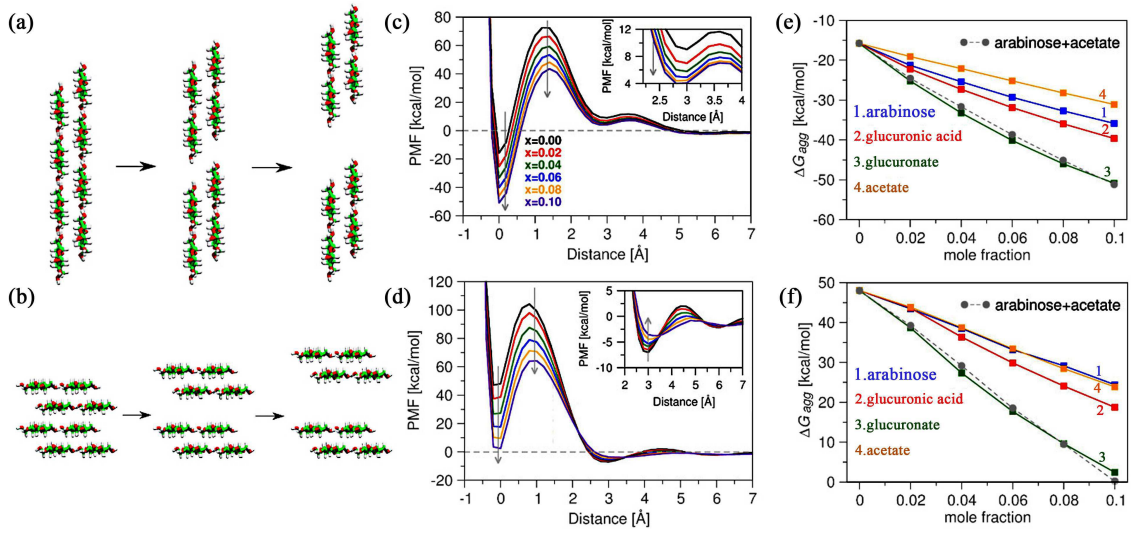


Figure 8. (Color online) Potential of mean force (PMF) and aggregation free energy ΔG_{agg} of cellulose nanocrystallites (CNs) in hemicellulose hydrogel. Parts (a), (b): Disaggregation of CNs by disrupting hydrophilic and hydrophobic contacts, respectively. Parts (c), (d): PMFs along disaggregation pathways a and b, respectively, at glucuronate molar fractions $x = 0.0 - 0.1$ [legend in part (c)]. Grey arrows indicate PMF change with glucuronate concentration. Parts (e) and (f): ΔG_{agg} for the hydrophilic and hydrophobic contacts (a) and (b), respectively, in hemicellulose hydrogels. Grey dotted line is the sum of the arabinose and acetate curves.

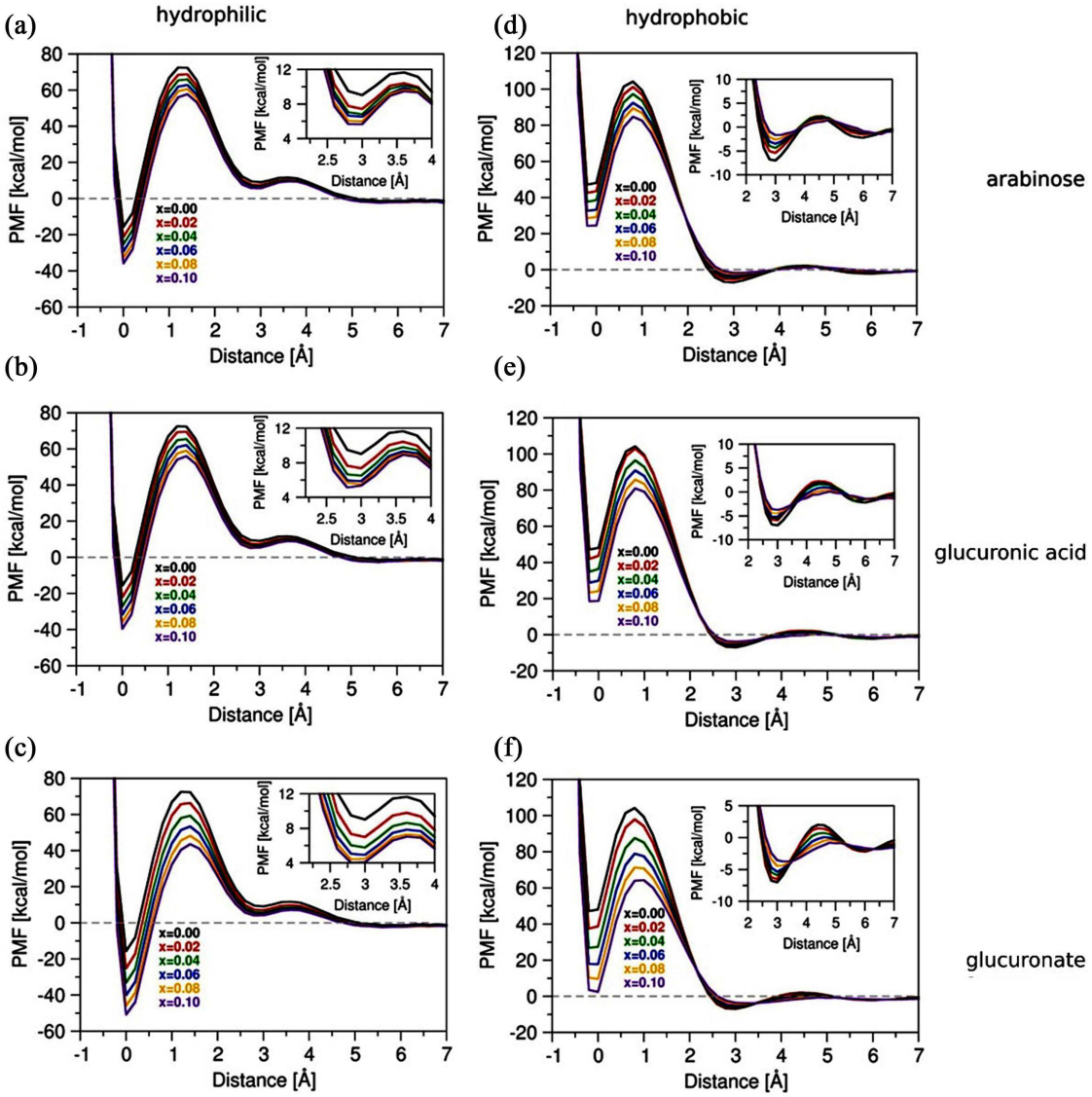


Figure 9. (Color online) Potential of mean force (PMF) for separation of the cellulose nanocrystallites along the hydrophilic and hydrophobic pathways is aqueous solution with different hemicellulose branches at different concentrations. PMFs along the hydrophilic pathway in (a) arabinose, (b) glucuronic acid, (c) glucuronate; along the hydrophobic pathway in (d) arabinose, (e) glucuronic acid, (f) glucuronate.

cellulose binding is controlled mainly by the site-specific H-bonds. The 3D-SFEDs for arabinose and glucuronic acid differ quantitatively, following the trends for the PMF and ΔG_{agg} in figures 8 and 9.

4. Conclusion

Based on the first principles foundation of statistical mechanics and Ornstein-Zernike type integral equation theory of molecular liquids, the three-dimensional reference interaction site model with the Kovalenko-Hirata closure relation (3D-RISM-KH molecular theory of solvation) [9–14] constitutes an essential component of multiscale modelling of chemical and biomolecular nanosystems in solution. As distinct from molecular simulations which explore the phase space of a molecular system by direct sampling, the molecular theory of solvation operates with spatial distributions rather than trajectories of molecules and is based on analytical summation of the free energy diagrams which yields the solvation

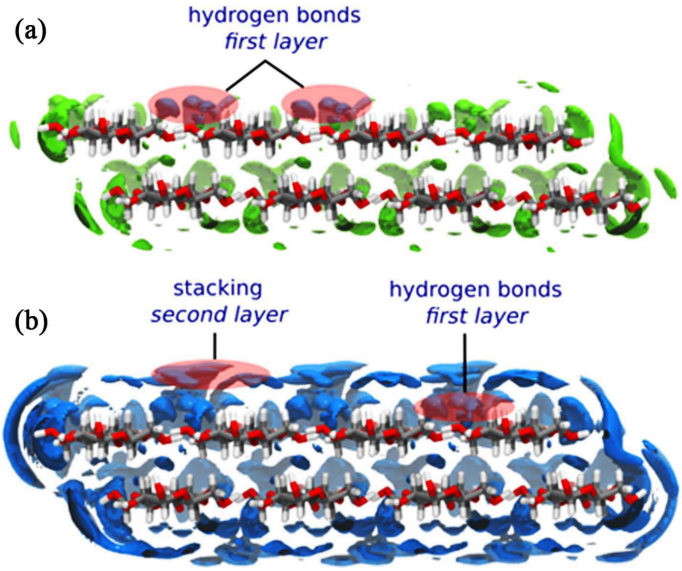


Figure 10. (Color online) Isosurfaces of the 3D solvation free energy density (3D-SFED) coming from glucuronate around the cellulose nanofibril. Part (a): The larger negative isovalue coming from highly localized distributions corresponds to hydrogen bonding. Part (b): The smaller negative isovalue coming from a diffuse second layer corresponding to hemicellulose-cellulose stacking.

structure and thermodynamics in the statistical-mechanical ensemble. It provides the solvation structure by solving the integral equations for the correlation functions and then the solvation thermodynamics analytically as a single integral of a closed form in terms of the correlation functions obtained.

The 3D-RISM-KH theory yields the solvation structure in terms of 3D maps of density distribution functions of solvent interaction sites around a solute molecule with full and consistent account for the effects of chemical functionalities of all solution species. The solvation free energy and subsequent thermodynamics are then obtained at once as a simple integral of the correlation functions by performing the thermodynamic integration analytically. Moreover, the possibility of analytical differentiation of the free energy functional enables: (i) self-consistent field coupling of 3D-RISM-KH with both *ab initio* type and density functional theory (DFT) quantum chemistry methods for multiscale description of electronic structure in solution [9, 12, 14–16], (ii) using 3D maps of potentials of mean force as scoring functions for molecular recognition [41, 56] and protein-ligand binding in docking protocols for fragment based drug design [41, 42, 57, 58], and (iii) hybrid molecular dynamics simulations performing quasidynamics of a biomolecule steered with 3D-RISM-KH mean solvation forces [17, 36–39].

Based on the first principles of statistical mechanics, the 3D-RISM-KH theory consistently reproduces, at the level of **fully converged** molecular simulation, the solvation structure and mean solvation forces in complex chemical and biomolecular nanosystems: both electrostatic forces (hydrogen bonding, other association, salt bridges, dielectric and Debye screening, ion localization) and nonpolar solvation effects (desolvation, hydrophobic hydration, hydrophobic interaction), as well as subtle interplays of these, such as preferential solvation, molecular recognition and ligand binding. This is very distinct from the continuum solvation schemes such as the Poisson-Boltzmann (PB) and Generalized Born (GB) models combined with the solvent accessible surface area (SASA) empirical nonpolar terms and additional volume and dispersion integral corrections, which are parameterized for hydration free energy of biomolecules but are neither really applicable to solvation structure effects in complex confined geometries nor transferable to solvent systems with cosolvent or electrolyte solutions at physiological concentrations.

For simple and complex solvents and solutions of a given composition, including buffers, salts, polymers, ligands and other cofactors at a finite concentration, the 3D-RISM-KH molecular theory of solvation properly accounts for chemical functionalities by representing in a single formalism both electrostatic and non-polar features of solvation, such as hydrogen bonding, structural solvent molecules, salt bridges, solvophobicity, and other electrochemical, associative and steric effects. For real systems, solving the 3D-

RISM-KH integral equations is far less computationally expensive than running molecular simulations which should be long enough to sample relevant exchange and binding events. This enables us to handle complex systems and processes occurring on large space and long time scales, problematic and frequently not feasible for molecular simulations. The 3D-RISM-KH theory has been validated on both simple and complex associating liquids with different chemical functionalities in a wide range of thermodynamic conditions, at different solid-liquid interfaces, in soft matter, and in various environments and confinements. The 3D-RISM-KH theory offers a “mental microscope” capable of providing insights into structure and molecular mechanisms of formation and functioning of various chemical [9–32] and biomolecular [14, 36–61] systems and synthetic organic [30–32] and biomass derived [33–35] nanomaterials.

Acknowledgements

The work was supported by the National Institute for Nanotechnology, National Research Council of Canada, University of Alberta, Natural Sciences and Engineering Research Council of Canada, Alberta Prion Research Institute, Alberta Innovates Bio Solutions, Alberta Innovates Technology Futures, ArboraNano – the Canadian Forest NanoProducts Network, Research Foundation of the State of Säo Paulo (FAPESP), and Brazilian Federal Agency for the Support and Evaluation of Graduate Education (CAPES). Computations were carried out on the high performance computing resources provided by the WestGrid – Compute/Calcul Canada national advanced computing platform. The author is grateful to Dr. Nikolay Blinov, Dr. Sergey Gusarov, Dr. Stanislav Stoyanov, Dr. Rodrigo Silveira, Prof. Leonardo Costa, Prof. Munir Skaf, Prof. Tyler Luchko, and Prof. Dr. Ihor Omelyan for their passionate contributions and ongoing interest and participation in these works.

References

1. Feynman R.P., In: Engineering and Science, Vol. 23, Caltech, Pasadena, 1960, 22–36.
2. Lundstrom M., Cummings P., Alam M., In: Nanotechnology Research Directions for Societal Needs in 2020: Retrospective and Outlook, Science Policy Reports Series Vol. 1, Roco M.C., Mirkin C.A., Hersam M.C. (Eds.), World Technology Evaluation Center Springer, Dordrecht, New York, 2011, chapter 1, 29–69.
3. Hansen J.-P., McDonald I., Theory of Simple Liquids, 3rd Edn., Elsevier Academic Press, London Burlington, 2006.
4. Hirata F., In: Molecular Theory of Solvation, Hirata F. (Ed.), Understanding Chemical Reactivity Series Vol. 24, Kluwer, Dordrecht, 2003, chapter 1, 1–60.
5. Chandler D., McCoy J.D., Singer S.J., J. Chem. Phys., 1986, **85**, 5971; doi:10.1063/1.451510.
6. Chandler D., McCoy J.D., Singer S.J., J. Chem. Phys., 1986, **85**, 5977; doi:10.1063/1.451511.
7. Beglov D., Roux B., J. Phys. Chem. B, 1997, **101**, 7821; doi:10.1021/jp971083h.
8. Kovalenko A., Hirata F., Chem. Phys. Lett., 1998, **290**, 237; doi:10.1016/S0009-2614(98)00471-0.
9. Kovalenko A., Hirata F., J. Chem. Phys., 1999, **110**, 10095; doi:10.1063/1.478883.
10. Kovalenko A., Hirata F., J. Chem. Phys., 2000, **112**, 10391; doi:10.1063/1.481676.
11. Kovalenko A., Hirata F., J. Chem. Phys., 2000, **112**, 10403; doi:10.1063/1.481677.
12. Kovalenko A., In: Molecular Theory of Solvation, Hirata F. (Ed.), Understanding Chemical Reactivity Series Vol. 24, Kluwer Academic Publishers, Norwell, 2003, chapter 4, 169–275.
13. Gusarov S., Pujari B.S., Kovalenko A., J. Comput. Chem., 2012, **33**, 1478; doi:10.1002/jcc.22974.
14. Kovalenko A., Pure Appl. Chem., 2013, **85**, No. 1, 159; doi:10.1351/PAC-CON-12-06-03.
15. Gusarov S., Ziegler T., Kovalenko A., J. Phys. Chem. A, 2006, **110**, 6083; doi:10.1021/jp054344t.
16. Casanova D., Gusarov S., Kovalenko A., Ziegler T., J. Chem. Theory Comput., 2007, **3**, 458; doi:10.1021/ct6001785.
17. Miyata T., Hirata F., J. Comput. Chem., 2008, **29**, 871; doi:10.1002/jcc.20844.
18. Kaminski J.W., Gusarov S., Wesolowski T.A., Kovalenko A., J. Phys. Chem. A, 2010, **114**, 6082; doi:10.1021/jp100158h.
19. Malvaldi M., Bruzzone S., Chiappe C., Gusarov S., Kovalenko A., J. Phys. Chem. B, 2009, **113**, 3536; doi:10.1021/jp810887z.
20. Kovalenko A., Kobryn A.E., Gusarov S., Lyubimova O., Liu X., Blinov N., Yoshida M., Soft Matter, 2012, **8**, 1508; doi:10.1039/C1SM06542D.
21. Kovalenko A., Hirata F., Chem. Phys. Lett., 2001, **349**, 496; doi:10.1016/S0009-2614(01)01241-6.
22. Kovalenko A., Hirata F., J. Theor. Comput. Chem., 2002, **01**, 381; doi:10.1142/S0219633602000282.

23. Yoshida K., Yamaguchi T., Kovalenko A., Hirata F., J. Phys. Chem. B, 2002, **106**, 5042; doi:10.1021/jp013400x.
24. Omelyan I., Kovalenko A., Hirata F., J. Theor. Comput. Chem., 2003, **2**, 193; doi:10.1142/S0219633603000501.
25. Huang W.J., Blinov N., Kovalenko A., J. Phys. Chem. B, 2015, **119**, 5588; doi:10.1021/acs.jpcb.5b01291.
26. Shapovalov V., Truong T.N., Kovalenko A., Hirata F., Chem. Phys. Lett., 2000, **320**, 186; doi:10.1016/S0009-2614(00)00191-3.
27. Stoyanov S.R., Gusarov S., Kovalenko A., In: Industrial Applications of Molecular Simulations, Meunier M. (Ed.), CRC Press, Boca Raton, 2012, chapter 14, 203–230.
28. Fafard J., Lyubimova O., Stoyanov S.R., Dedzo G.K., Gusarov S., Kovalenko A., Detellier C., J. Phys. Chem. C, 2013, **117**, 18556; doi:10.1021/jp4064142.
29. Huang W.-J., Dedzo G.K., Stoyanov S.R., Lyubimova O., Gusarov S., Singh S., Lao H., Kovalenko A., Detellier C., J. Phys. Chem. C, 2014, **118**, 23821; doi:10.1021/jp507393u.
30. Moralez J.G., Raez J., Yamazaki T., Motkuri R.K., Kovalenko A., Fenniri H., J. Am. Chem. Soc., 2005, **127**, 8307; doi:10.1021/ja051496t.
31. Johnson R.S., Yamazaki T., Kovalenko A., Fenniri H., J. Am. Chem. Soc., 2007, **129**, 5735; doi:10.1021/ja0706192.
32. Yamazaki T., Fenniri H., Kovalenko A., ChemPhysChem, 2010, **11**, 361; doi:10.1002/cphc.200900324.
33. Silveira R.L., Stoyanov S.R., Gusarov S., Skaf M.S., Kovalenko A., J. Am. Chem. Soc., 2013, **135**, 19048; doi:10.1021/ja405634k.
34. Kovalenko A., Nord. Pulp Pap. Res. J., 2014, **29**, 144; doi:10.3183/NPPRJ-2014-29-01-p144-155.
35. Silveira R.L., Stoyanov S.R., Gusarov S., Skaf M.S., Kovalenko A., J. Phys. Chem. Lett., 2015, **6**, 206; doi:10.1021/jz502298q.
36. Luchko T., Gusarov S., Roe D.R., Simmerling C., Case D.A., Tuszynski J., Kovalenko A., J. Chem. Theory Comput., 2010, **6**, 607; doi:10.1021/ct900460m.
37. Omelyan I., Kovalenko A., Mol. Simul., 2013, **39**, 25; doi:10.1080/08927022.2012.700486.
38. Omelyan I., Kovalenko A., J. Chem. Phys., 2013, **139**, 244106; doi:10.1063/1.4848716.
39. Omelyan I.P., Kovalenko A., J. Chem. Theory Comput., 2015, **11**, 1875; doi:10.1021/ct5010438.
40. Yoshida N., Phongphanphanee S., Maruyama Y., Imai T., Hirata F., J. Am. Chem. Soc., 2006, **128**, 12042; doi:10.1021/ja0633262.
41. Yoshida N., Imai T., Phongphanphanee S., Kovalenko A., Hirata F., J. Phys. Chem. B, 2009, **113**, 873; doi:10.1021/jp807068k.
42. Imai T., Oda K., Kovalenko A., Hirata F., Kidera A., J. Am. Chem. Soc., 2009, **131**, 12430; doi:10.1021/ja905029t.
43. Phongphanphanee S., Yoshida N., Hirata F., J. Am. Chem. Soc., 2008, **130**, 1540; doi:10.1021/ja077087+.
44. Phongphanphanee S., Rungrotmongkol T., Yoshida N., Hannongbua S., Hirata F., J. Am. Chem. Soc., 2010, **132**, 9782; doi:10.1021/ja1027293.
45. Maruyama Y., Yoshida N., Hirata F., Interdiscip. Sci. Comput. Life Sci., 2011, **3**, 290; doi:10.1007/s12539-011-0104-7.
46. Yonetani Y., Maruyama Y., Hirata F., Kono H., J. Chem. Phys., 2008, **128**, 185102; doi:10.1063/1.2904865.
47. Maruyama Y., Yoshida N., Hirata F., J. Phys. Chem. B, 2010, **114**, 6464; doi:10.1021/jp912141u.
48. Harano Y., Imai T., Kovalenko A., Kinoshita M., Hirata F., J. Chem. Phys., 2001, **114**, 9506; doi:10.1063/1.1369138.
49. Imai T., Takekiyo T., Kovalenko A., Hirata F., Kato M., Taniguchi Y., Biopolymers, 2005, **79**, 97; doi:10.1002/bip.20337.
50. Imai T., Kinoshita M., Hirata F., J. Chem. Phys., 2000, **112**, 9469; doi:10.1063/1.481565.
51. Imai T., Ohyama S., Kovalenko A., Hirata F., Protein Sci., 2007, **16**, 1927; doi:10.1110/ps.072909007.
52. Kovalenko A., In: Volume Properties: Liquids, Solutions and Vapours, Wilhelm E., Letcher T. (Eds.), Royal Society of Chemistry, Cambridge, 2015, chapter 22, 575–610; doi:10.1039/9781782627043-00575.
53. Blinov N., Dorosh L., Wishart D., Kovalenko A., Biophys. J., 2010, **98**, 282; doi:10.1016/j.bpj.2009.09.062.
54. Yamazaki T., Blinov N., Wishart D., Kovalenko A., Biophys. J., 2008, **95**, 4540; doi:10.1529/biophysj.107.123000.
55. Blinov N., Dorosh L., Wishart D., Kovalenko A., Mol. Simul., 2011, **37**, 718; doi:10.1080/08927022.2010.544306.
56. Genheden S., Luchko T., Gusarov S., Kovalenko A., Ryde U., J. Phys. Chem. B, 2010, **114**, 8505; doi:10.1021/jp101461s.
57. Kovalenko A., Blinov N., J. Mol. Liq., 2011, **164**, 101; doi:10.1016/j.molliq.2011.09.011.
58. Nikolić D., Blinov N., Wishart D., Kovalenko A., J. Chem. Theory Comput., 2012, **8**, 3356; doi:10.1021/ct300257v.
59. Imai T., Miyashita N., Sugita Y., Kovalenko A., Hirata F., Kidera A., J. Phys. Chem. B, 2011, **115**, 8288; doi:10.1021/jp2015758.
60. Huang W.-J., Blinov N., Wishart D.S., Kovalenko A., J. Chem. Inform. Modeling, 2015, **55**, 317; doi:10.1021/ci500520q.
61. Stumpe M.C., Blinov N., Wishart D., Kovalenko A., Pande V.S., J. Phys. Chem. B, 2011, **115**, 319; doi:10.1021/jp102587q.
62. Perkyns J.S., Lynch G.C., Howard J.J., Pettitt B.M., J. Chem. Phys., 2010, **132**, 064106; doi:10.1063/1.3299277.

63. Kast S.M., Kloss T., J. Chem. Phys., 2008, **129**, 236101; doi:10.1063/1.3041709.
64. Joung I.S., Luchko T., Case D.A., J. Chem. Phys., 2013, **138**, 044103; doi:10.1063/1.4775743.
65. Perkyns J.S., Pettitt B.M., Chem. Phys. Lett., 1992, **190**, 626; doi:10.1016/0009-2614(92)85201-K.
66. Perkyns J.S., Pettitt B.M., J. Chem. Phys., 1992, **97**, 7656; doi:10.1063/1.463485.
67. Kvamme B., Int. J. Thermophys., 1995, **16**, 743; doi:10.1007/BF01438859.
68. Free Energy Calculations: Theory and Applications in Chemistry and Biology, Springer Series in Chemical Physics Vol. 86, Chipot C., Pohorille A. (Eds.), Springer-Verlag, Berlin, Heidelberg, 2007; doi:10.1007/978-3-540-38448-9.
69. Yamazaki T., Kovalenko A., J. Chem. Theory Comput., 2009, **5**, 1723; doi:10.1021/ct9000729.
70. Yamazaki T., Kovalenko A., J. Phys. Chem. B, 2011, **115**, 310; doi:10.1021/jp1082938.
71. Yu H.-A., Karplus M., J. Chem. Phys., 1988, **89**, 2366; doi:10.1063/1.455080.
72. Yu H.-A., Roux B., Karplus M., J. Chem. Phys., 1990, **92**, 5020; doi:10.1063/1.458538.
73. Chandler D., Singh Y., Richardson D.M., J. Chem. Phys., 1984, **81**, 1975; doi:10.1063/1.447820.
74. Chandler D.. Phys. Rev. E, 1993, **48**, 2898; doi:10.1103/PhysRevE.48.2898.
75. Ichiye T., Chandler D., J. Phys. Chem., 1988, **92**, 5257; doi:10.1021/j100329a037.
76. Lee P.H., Maggiora G.M., J. Phys. Chem., 1993, **97**, 10175; doi:10.1021/j100141a045.
77. Palmer D.S., Frolov A.I., Ratkova E.L., Fedorov M.V., J. Phys.: Condens. Matter, 2010, **22**, 492101; doi:10.1088/0953-8984/22/49/492101.
78. Truchon J.-F., Pettitt B.M., Labute P., J. Chem. Theory Comput., 2014, **10**, 934; doi:10.1021/ct4009359.
79. Ng K.-C., J. Chem. Phys., 1974, **61**, 2680; doi:10.1063/1.1682399.
80. Kovalenko A., Ten-no S., Hirata F., J. Comput. Chem., 1999, **20**, 928; doi:10.1002/(SICI)1096-987X(19990715)20:9<928::AID-JCC4>3.0.CO;2-X.
81. Pulay P., Chem. Phys. Lett., 1980, **73**, 393; doi:10.1016/0009-2614(80)80396-4.
82. Saad Y., Schultz M.H., SIAM J. Sci. Stat. Comput., 1986, **7**, 856; doi:10.1137/0907058.
83. Schmeer G., Maurer A., Phys. Chem. Chem. Phys., 2010, **12**, 2407; doi:10.1039/b917653e.
84. Bowron D.T., Finney J.L., Soper A.K., J. Phys. Chem. B, 1998, **102**, 3551; doi:10.1021/jp972780c.
85. Bowron D.T., Finney J.L., Soper A.K., Molec. Phys., 1998, **93**, 531; doi:10.1080/002689798168871.
86. Sangster J., J. Phys. Chem. Ref. Data, 1989, **18**, 1111; doi:10.1063/1.555833.
87. Sangster J., Octanol-Water Partition Coefficients: Fundamentals and Physical Chemistry, Wiley, Chichester, 1997.
88. Van de Waterbeemd H., Camenisch G., Folkers G., Chretien J.R., Raevsky O.A., J. Drug Targeting, 1998, **6**, 151; doi:10.3109/10611869808997889.
89. Li J., Zhu T., Hawkins G.D., Winget P., Liotard D.A., Cramer C.J., Truhlar D.G., Theor. Chem. Acc., 1999, **103**, 9; doi:10.1007/s002140050513.
90. Wang J., Wang W., Huo S., Lee M., Kollman P.A., J. Phys. Chem. B, 2001, **105**, 5055; doi:10.1021/jp0102318.
91. Himmel M.E., Ding S.-Y., Johnson D.K., Adney W.S., Nimlos M.R., Brady J.W., Foust T.D., Science, 2007, **315**, 804; doi:10.1126/science.1137016.
92. Chundawat S.P.S., Beckham G.T., Himmel M.E., Dale B.E., Ann. Rev. Chem. Biomolec. Eng., 2011, **2**, 121; doi:10.1146/annurev-chembioeng-061010-114205.
93. Mortimer J.C., Miles G.P., Brown D.M., Zhang Z., Segura M.P., Weimar T., Yu X., Seffen K.A., Stephens E., Turner S.R., Dupree P., Proc. Natl. Acad. Sci. USA, 2010, **107**, 17409; doi:10.1073/pnas.1005456107.
94. Ding S.-Y., Liu Y.-S., Zeng Y., Himmel M.E., Baker J.O., Bayer E.A., Science, 2012, **338**, 1055; doi:10.1126/science.1227491.
95. DeMartini J.D., Pattathil S., Miller J.S., Li H., Hahn M.G., Wyman C.E., Energy Environ. Sci., 2013, **6**, 898; doi:10.1039/c3ee23801f.
96. Beckham G.T., Matthews J.F., Peters B., Bomble Y.J., Himmel M.E., Crowley M.F., J. Phys. Chem. B, 2011, **115**, 4118; doi:10.1021/jp1106394.
97. Payne C.M., Himmel M.E., Crowley M.F., Beckham G.T., J. Phys. Chem. Lett., 2011, **2**, 1546; doi:10.1021/jz2005122.
98. Lindner B., Petridis L., Schulz R., Smith J.C., Biomacromolecules, 2013, **14**, 3390; doi:10.1021/bm400442n.
99. Petridis L., Schulz R., Smith J.C., J. Am. Chem. Soc., 2011, **133**, 20277; doi:10.1021/ja206839u.
100. Pauly M., Keegstra K., Plant J., 2008, **54**, 559; doi:10.1111/j.1365-313X.2008.03463.x.
101. Gomes T.C.F., Skaf M.S., J. Comput. Chem., 2012, **33**, 1338; doi:10.1002/jcc.22959.

Молекулярна теорія сольватації: методологічний підсумок та ілюстрації

А. Коваленко[1,2]

[1] Національний інститут нанотехнологій, 11421 Саскачеван Драйв, Едмонтон, AB, T6G 2M9, Канада

[2] Факультет машинобудування, Альбертський університет, Едмонтон, AB, T6G 2G7, Канада

Теорія інтегральних рівнянь, яка базується на принципах статистичної механіки, є багатообіцяючою з огляду на її місце в різномасштабній методології для розчинів хімічних та бімолекулярних наносистем. Стартуючи з міжмолекулярних потенціалів взаємодії, ця методологія використовує діаграмний аналіз вільної енергії сольватації в стистично-механічному ансамблі. Застосування відповідних умов замикання дає можливість звести нескінчений ланцюжок зв'язаних інтегральних рівнянь для багаточастинкових кореляційних функцій до системи рівнянь для дво- або трьохчастинкових кореляційних функцій. Розв'язок цих рівнянь дає результати для сольватаційної структури, точність яких є співмірною з результатами комп'ютерного моделювання, але має перевагу стосовно трактування ефектів і процесів, пов'язаних з великими відстанями та довгими часами, які якраз і створюють проблеми для комп'ютерного моделювання при явному врахуванні розчинника. Одна з версій цієї методології, теорія інтегральних рівнянь тривимірної моделі взаємодіючих силових центрів (3D-RISM) з умовами замикання Коваленка-Хірати (KH), дозволяє отримати 3D мапу кореляційних функцій, включаючи розподіл густини взаємодіючих силових центрів розчинника навколо супрамолекули, з повністю самоузгодженим врахуванням ефектів хімічної функціональності всіх складників розчину. Вільна енергія сольватації та відповідна їй термодинаміка природним чином отримується як інтеграл від 3D кореляційних функцій з тим, що термодинамічне інтегрування виконується аналітично. Аналітична форма функціоналу вільної енергії дозволяє самоузгоджене поєднання 3D-RISM-KH теорії з методами квантової хімії при різномасштабному описі електронної структури розчину, використання 3D мапи потенціалів середньої сили як рейтингової функції для молекулярного розпізнавання та білок-ліганд зв'язування в докових протоколах при фрагментарній розробці медичних препаратів, а також при гібридному молекулярно-динамічному моделюванні квазідинаміки біомолекул на основі 3D-RISM-KH потенціалів середньої сили. 3D-RISM-KH теорія була протестована як на простих, так і на складних асоційованих рідинах з різними хімічними зв'язками та в широкому діапазоні термодинамічних параметрів, на контакті з твердими та м'якими поверхнями різної природи та в різних середовищах. 3D-RISM-KH теорія є таким собі "інтелектуальним мікроскопом", який здатний демонструвати структуру та механізм формування і функціонування різних хімічних та біологічних систем і наноматеріалів.

Ключові слова: *хімія розчинів, бімолекулярна сольватація, метод інтегральних рівнянь в теорії рідини, 3D-RISM-KH молекулярна теорія сольватації, сольватаційна структура і термодинаміка, потенціал середньої сили*