3D Molecular Theory of Solvation Coupled with MD for Nanomedical Sciences

Andriy Kovalenko[*,**]

[*]National Institute for Nanotechnology
11421 Saskatchewan Dr., Edmonton, AB, T6G 2M9, Canada, andriy.kovalenko@nrc.ca
[**]Department of Mechanical Engineering, University of Alberta, Edmonton, AB, T6G 2G8, Canada

ABSTRACT

We coupled statistical-mechanical, molecular theory of solvation (a.k.a. 3D-RISM-KH) contracting solvent degrees of freedom with MD simulation in the Amber package. This included a number of accelerating schemes with cutoffs and iterative guess for the correlation functions, and extrapolating solvent-induced forces and applying them in large multi-time steps (up to 20 fs) to enable simulation of large biomolecules. The MD/3D-RISM-KH method allows one to study biomolecular processes on extremely long timescales, as the statistics of rare solvent and ligand events is accounted for analytically. It replaces MM/GB(PB)SA post-processing using empirical treatment of non-polar contributions with MM/3D-RISM-KH evaluation of the solvation thermodynamics. The method accurately yields the solvation structure for biomolecular systems as large as a chaperon (GroEL) and 3D maps of ligand binding affinity at once without phenomenological approximations, and has significant potential for computer-aided drug design.

Keywords: 3D molecular theory of solvation, molecular dynamics, AMBER package, ligand binding affinity, chaperon, ion channel, prion proteins

1 INTRODUCTION

Molecular dynamics (MD) simulation with explicit solvent yields detailed modeling of solvated biomolecules for processes within accessible time scales. A major computational burden comes from solvent molecules. Moreover, solvent enters pockets and inner cavities of proteins through a very slow process of their conformational changes, nearly as difficult to model as protein folding. The generalized Born model represents the solvent polarization effects by a cavity in dielectric continuum (optionally, with Debye screening), whereas the non-electrostatic contributions are empirically parameterized. However, it bears the fundamental drawbacks of implicit solvation: hydrogen bonding and hydrophobic interactions are non-transferable to new complex systems (e.g. cosolvent or different buffer ions); the solvent accessible surface and volumetrics are not well defined, the entropic term is absent in continuum solvation. An alternative is statistical-mechanical, molecular theory of solvation, a.k.a. three-dimensional reference interaction site model, which starts from an explicit solvent model with all-atom force field but operates with 3D correlation functions of species in a statistical ensemble rather than with individual molecular trajectories and predicts the solvation structure and thermodynamics of biomolecules from the first principles of statistical mechanics [1]. Complemented with the Kovalenko-Hirata (KH) closure approximation [1], the 3D-RISM-KH theory properly accounts for chemical functionalities of both biomolecules and solvent by representing both electrostatic and non-polar features of the solvation structure, such as hydrogen bonding, hydrophobicity, salt bridges, structural solvent, etc; moreover, it analytically yields the solvation thermodynamics, including the solvation free energy, its energetic and entropic decomposition, and partial molar volume effects [1-5]. Recently, we have coupled the 3D-RISM-KH method contracting solvent degrees of freedom with molecular dynamics simulation of biomolecules in the Amber molecular dynamics package [6]. This included a number of accelerating schemes with several cutoffs for the interaction potentials and correlation functions, an iterative guess for the 3D-RISM solutions, and extrapolating solvent-induced forces and applying them in large multi-time steps (up to 20 fs) to enable simulation of large biomolecules. The coupled MD/3D-RISM-KH method makes feasible modeling of biomolecular structures of practical interest and thus has tremendous potential for computer-aided drug design. It allows one to study biomolecular systems involving solvent processes occurring on extremely long timescales, as the statistics of rare solvent events such as exchange of solvent molecules and ions, and ligand binding at pockets and inner spaces is accounted for statistically-mechanically. The method replaces the MM/GBSA or MM/PBSA post-processing suffering from the empirical treatment of non-polar contributions with the MM/3D-RISM-KH accurate evaluation of the solvation thermodynamics [6,7]. It also addresses the problem of molecular recognition [1,8,9] by yielding at once 3D maps of binding affinity without any phenomenological approximations [8,9], which can be used to improve the quality of 3D site binding maps used in fragment-based screening of drug compounds in most of the novel biomolecular software packages. Below we briefly summarize the 3D molecular theory of solvation and show how it yields the solvation structure related to functions of biomolecular systems as large and complex as a Gloeobacter violaceus pentameric ligand-gated ion channel (GLIC) and a solvated chaperon (GroEL) [10], and how it predicts binding maps of prion proteins for development of novel inhibitors of prion protein conversion [11].

2 THREE-DIMENSIONAL MOLECULAR THEORY OF SOLVATION

The solvation structure is represented by the probability density of finding site γ of solvent molecules at 3D space position **r** around the solute (supra)molecule, $\rho_\gamma g_\gamma(\mathbf{r})$, which is determined by the average number density ρ_γ in the solution bulk times the 3D distribution function (normalized density distribution) $g_\gamma(\mathbf{r})$ of solvent site γ. The latter indicates site density enhancement when $g_\gamma(\mathbf{r})>1$ or depletion when $g_\gamma(\mathbf{r})<1$ 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 are obtained from the 3D-RISM integral equation [1]

$$h_\gamma(\mathbf{r}) = \sum_\alpha \int d\mathbf{r}' c_\alpha(\mathbf{r}-\mathbf{r}')\chi_{\alpha\gamma}(r'), \quad (1)$$

where $h_\gamma(\mathbf{r})$ is the 3D total correlation function of solvent site γ related to the 3D site distribution function by $g_\gamma(\mathbf{r}) = h_\gamma(\mathbf{r}) + 1$, and $c_\gamma(\mathbf{r})$ is the 3D direct correlation function which has the asymptotics of the solute-solvent site interaction potential, $c_\gamma(\mathbf{r}) \sim - u_\gamma(\mathbf{r}) / (k_B T)$; the site-site susceptibility of pure solvent $\chi_{\alpha\gamma}(r)$ is an input to the 3D-RISM theory; and indices α and γ enumerate all sites on all sorts of solvent species. Another relation between the 3D total and direct correlation functions, called a closure, is necessary to complement the 3D-RISM integral equation (1). The exact closure can be formally expressed as a series in terms of multiple integrals of the combinations of the total correlation functions. However, it is extremely cumbersome, and in practice is replaced with amenable approximations. We use the 3D-KH closure approximation, proven to be appropriate to describe various association effects in complex liquids and electrolyte solutions [1], and in supramolecular synthetic organic [2-4] and biomolecular [5-11] systems in solution,

$$g_\gamma(\mathbf{r}) = \begin{cases} \exp(d_\gamma(\mathbf{r})) & \text{for} \quad d_\gamma(\mathbf{r}) \le 0 \\ 1 + d_\gamma(\mathbf{r}) & \text{for} \quad d_\gamma(\mathbf{r}) > 0 \end{cases}, \quad (2)$$

$$d_\gamma(\mathbf{r}) = -\frac{u_\gamma(\mathbf{r})}{k_B T} + h_\gamma(\mathbf{r}) - c_\gamma(\mathbf{r}),$$

where $u_\gamma(\mathbf{r})$ is the 3D interaction potential between the whole solute and solvent site γ specified by the molecular force field, and $k_B T$ is the Boltzmann constant times the solution temperature. The closure (3) couples in a nontrivial way the so-called mean spherical approximation (MSA) and hypernetted chain (HNC) approximation, the former being applied to the spatial regions of solvent density enrichment ($g_\gamma(\mathbf{r})>1$) such as association peaks, and the latter to those of solvent density depletion ($g_\gamma(\mathbf{r})<1$) including the repulsive core [1].

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

$$\chi_{\alpha\gamma}(r) = \omega_{\alpha\gamma}(r) + \rho_\alpha h_{\alpha\gamma}(r), \quad (3)$$

where the intramolecular correlation function

$$\omega_{\alpha\gamma}(r) = \delta_{\alpha\gamma}\delta(r) + (1-\delta_{\alpha\gamma})\delta(r - l_{\alpha\gamma})/(4\pi l_{\alpha\gamma}^2)$$

represents the geometry of solvent molecules with site-site separations $l_{\alpha\gamma}$ specified by the molecular force field (z-matrix in quantum chemistry), and $h_{\alpha\gamma}(r)$ is the radial total correlation function between sites α and γ enumerating all sites on all sorts of molecules in bulk solvent. In advance to the 3D-RISM-KH calculation, $h_{\alpha\gamma}(r)$ are obtained from the dielectrically consistent RISM theory [12] coupled with the KH closure (DRISM-KH) [1], applied to the bulk solvent with counter ions, co-solvent, and ligands at a given concentration. The susceptibility (3) of the bulk solution is then input into the 3D-RISM integral equation (1).

The solvation free energy of the solute supramolecule in multicomponent solvent following from the 3D-RISM-KH integral equations (1) and (2) is given by the closed analytical expression [1]

$$\mu_{\text{solv}} = k_B T \sum_\gamma \rho_\gamma \int d\mathbf{r} \left[\frac{1}{2} h_\gamma^2(\mathbf{r}) \Theta(-h_\gamma(\mathbf{r})) - c_\gamma(\mathbf{r}) - \frac{1}{2} h_\gamma(\mathbf{r}) c_\gamma(\mathbf{r}) \right], \quad (4)$$

where $\Theta(x)$ is the Heaviside step function.

The potential of mean force $W_\gamma(\mathbf{r})$ acting on solvent site γ near the biomolecule is defined in terms of the 3D site distribution function as $W_\gamma(\mathbf{r}) = -k_B T \ln g_\gamma(\mathbf{r})$.

For a structural water molecule "w", the binding strength is determined as the difference between the potential of mean force of water oxygen O in the effective potential well at the biomolecule and that in the first peak of bulk water. The binding strength is thus expressed in terms of the water oxygen peaks at the biomolecule and in bulk water as

$$A_{\text{w}}^{\text{binding}} = W_{\text{O}}^{\text{max at solute}} - W_{\text{O}}^{\text{max in bulk}} = -k_B T \ln\left(g_{\text{O}}(\mathbf{r}^{\text{max at solute}}) / g_{\text{O}}(\mathbf{r}^{\text{max in bulk}}) \right). \quad (5)$$

The solvation free energy (4) and binding energy (5) obtained by 3D-RISM-KH can be decomposed into partial contributions of solute sites, providing a basis for spatial decomposition analysis (SDA) of association effects [13].

To properly treat electrostatic forces in electrolyte solution with polar molecular solvent and ionic species when evaluating the convolution in the 3D-RISM-KH integral equations (1)-(2), the radial correlations (3) of bulk solvent, and in the thermodynamics expressions (4) and (5), the electrostatic asymptotics of all the correlation functions (both the 3D and radial ones) are treated analytically [1,14]. The non-periodic electrostatic asymptotics are separated out in the direct and reciprocal space and the remaining short-range terms of the correlation functions are discretized on a

rectangular grid in a non-periodic uniform box large enough to ensure decay of the short-range terms at the box boundaries [14]. The convolution of the short-range terms in the integral equation (1) is calculated using 3D fast Fourier transform (3D-FFT) in double size box to prevent aliasing. Accordingly, the electrostatic asymptotics terms in the thermodynamics integral (4) are handled analytically, reducing to one-dimensional integrals easy to compute [14]. This analytical treatment of the electrostatics eliminates the periodicity artifacts arising when using the Ewald summation and 3D-FFT supercell for the electrostatic terms, in particular, those in the values of $g_\gamma(\mathbf{k}=0)$ coming from the cancellation or renormalization of the Coulomb singularities of $c_\gamma(\mathbf{k}\to 0)$ [1,14].

The equations (1)-(2) are converged to a relative root mean square accuracy (typically 10^{-5}) ensuring accurate thermodynamics, by using the modified direct inversion in the iterative subspace (MDIIS) accelerated numerical solver of integral equations of liquid state theory [1].

3 MD COUPLED WITH 3D-RISM-KH

In the coupled MD/3D-RISM-KH method implemented in Amber molecular dynamics package [6], MD is applied to the biomolecule while the solvation structure, free energy and associated forces are obtained by 3D-RISM-KH. The latter are derived by differentiating the Kirkwood thermodynamic integration formula analytically, also valid for the solvation free energy (4) in the 3D-KH approximation (2),

$$\mathbf{f}(\mathbf{R}_i) \equiv -\frac{\partial \mu_{\text{solv}}}{\partial \mathbf{R}_i} = \sum_\gamma \rho_\gamma \int d\mathbf{r}\, g_\gamma(\mathbf{r}) \frac{\partial u_{i\gamma}(\mathbf{r}-\mathbf{R}_i)}{\partial \mathbf{R}_i}, \quad (6)$$

where $u_{i\gamma}(\mathbf{r}\text{-}\mathbf{R}_i)$ is the pairwise interaction potential between biomolecule site i located at $\mathbf{R}_i$ and solvent site γ at $\mathbf{r}$. To obtain meaningful sampling of solute conformations, we reduce the computational expense of 3D-RISM calculations by several optimization strategies: (i) creating high-quality initial guesses to the direct correlation function $c_\gamma(\mathbf{r})$ from multiple previous solutions; (ii) accelerating the pre- and post-processing of the solute-solvent potentials, long-range asymptotics, and forces using a cutoff scheme and minimal solvation box; and (iii) avoiding direct calculation of 3D-RISM-KH solvation forces $\{\mathbf{F}\}^{(k)}$ at current time step k by interpolating them based of those from previous time steps, or force-coordinate extrapolation multi-time step procedure,

$$\{\mathbf{F}\}^{(k)} = \sum_{l=1}^{N} a_{kl} \{\mathbf{F}\}^{(l)}, \quad l \in 3\text{D-RISM steps}, \quad (7)$$

where the weight coefficients a_{kl} are obtained as the best representation of the matrix of arrangement of solute atoms $\{\mathbf{R}\}^{(k)}$ at current time step k in terms of its projections onto the "basis" of l=1,N previous ones $\{\mathbf{R}\}^{(l)}$ obtained from the 3D-RISM-KH, by minimizing the norm of the difference

$$\text{minimize} \left| \{\mathbf{R}\}^{(k)} - \sum_{l=1}^{N} a_{kl} \{\mathbf{R}\}^{(l)} \right|^2. \quad (8)$$

4 3D-RISM-KH PREDICTIONS FOR BIOMOLECULAR SOLVATION

The 3D-RISM-KH equations (1)-(2) were solved on a uniform 3D grid of 256×256×256 points in a cubic supercell of size 128 Å, large enough to accommodate the biomolecule together with sufficient solvation space. Further refinement of the grid does not affect the results much. The solute-solvent interaction potentials $u_{i\gamma}(\mathbf{r}\text{-}\mathbf{R}_i)$ comprised the Coulomb and Lennard-Jones terms with the potential parameters from the OPLS-AA force field, and the SPC/E model was used for ambient water solvent with the bulk density 0.99705 g/cm3 and dielectric constant 78.4 at the temperature T=300 K and pressure 1 bar.

We used the 3D-RISM-KH method to determine the most probable binding modes of 2-pyrrolidin-1-yl-N-[4-[4-(2-pyrrolidin-1-yl-acetylamino)-benzyl]-phenyl]-acetamide to mouse prion protein (PrP^c). The simulation produced the 3D distribution of each interaction site of the ligand around the protein, which then were processed to yield the most probable binding locations of the bound sites (Figure 1). This method predicted the hydrogen bonding between the ligand and protein in full agreement with experiment [11].

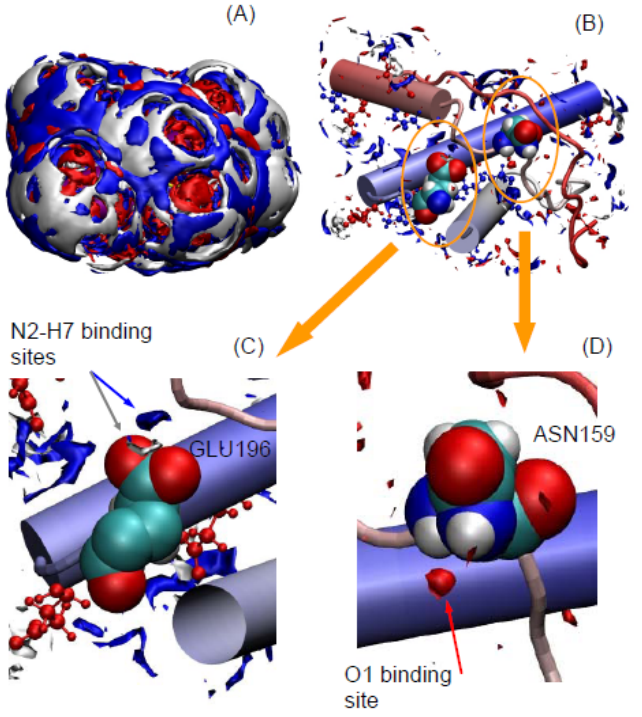

Figure 1: Predictions of the 3D-RISM-KH theory for ligand binding to prion protein [11]. (A): 3D density maps of the ligand O (red), N (blue), and H (white) sites in the first solvation shell. (B): Possible binding modes represented with high density pockets of ligand O, N, and H. Shown are data for hydrogen-bond forming atoms. Residues GLU196 and ASN159 are represented with van der Waals spheres. (C-D): Zoomed view of the N-H and O atomic binding sites in the proximity of the GLU196 and ASN159 residues.

We also validated the 3D-RISM-KH molecular theory of solvation against MD simulation for large biomolecular systems. The coordination numbers of water in the hydration shells in the internal cavity of GroEL chaperon protein are well correlated (up to 0.98) with the available MD results [10]. This opens up 3D-RISM-KH use to study biologically relevant properties related to solvation effects, while calculation of such equilibrium properties by conventional methods poses much computational challenge.

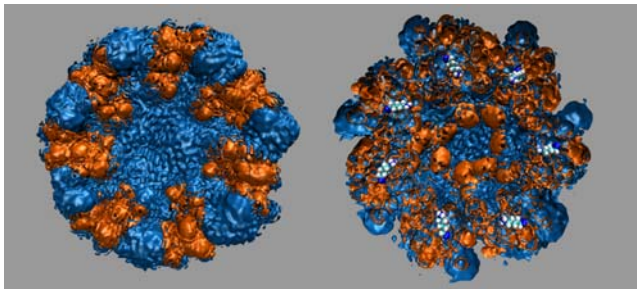

Figure 2: Predictions of the 3D-RISM-KH molecular theory of solvation for the 3D distributions of Na^+ (blue) and Cl^- (orange) ions in 0.2M aqueous electrolyte solution around the GroEL chaperon protein [10]. First solvation shell at the entrance (left panel) and inside the chaperon (right panel).

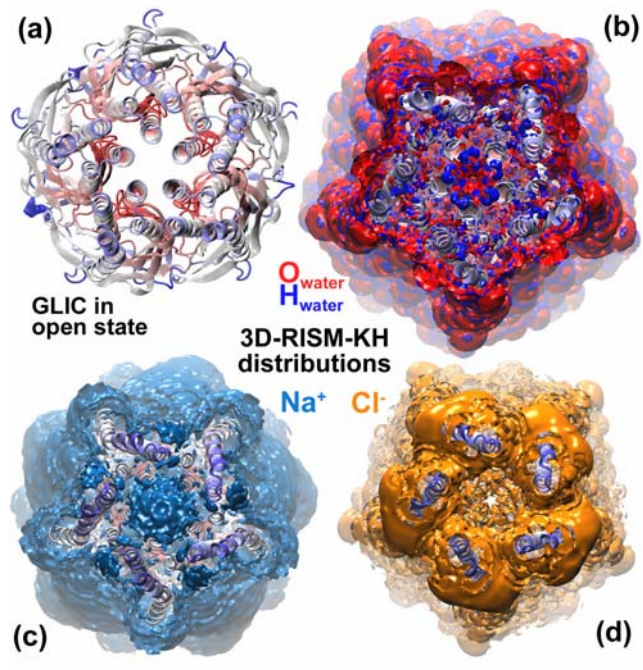

Figure 3: 3D solvent distributions in Gloeobacter violaceus pentameric ligand-gated ion channel (GLIC), predicted by the 3D-RISM-KH molecular theory of solvation. (a) GLIC in an open state (N. Bocquet *et al*, Nature, 457, 111, 2009); top view at the ring of hydrophilic residues of GLIC M2 helices. 3D solvent distributions: (b) water O (red) and H (blue); (c) Na^+ (turquoise); (d) Cl^- (orange).

For further illustration, Figure 2 shows the 3D distributions of ions around GroEL in NaCl aqueous solution.

Figure 3 exhibits the 3D-RISM-KH predictions for the 3D distributions of 0.2M NaCl aqueous solution around the ligand-gated ion channel (GLIC). The theory explicitly reproduces the solvation features related to the permeation properties of GLIC: Water forms a cap at the ring of hydrophilic residues inside GLIC, Na^+ ions penetrate and fill the channel, and Cl^- ions are depleted in the channel.

Acknowledgments. This work was supported by the Prion Core Project of the Alberta Prion Research Institute (APRI) and by the National Research Council (NRC) of Canada.

REFERENCES

[1] A. Kovalenko, Three-dimensional RISM theory for molecular liquids and solid-liquid interfaces, in: *Molecular Theory of Solvation*, F. Hirata (ed.). Series: *Understanding Chemical Reactivity*, vol.24, (Kluwer, Dordrecht, 2003) pp.169-275, (and references therein).

[2] J. G. Moralez, J. Raez, T. Yamazaki, R. K. Motkuri, A. Kovalenko, H Fenniri, J. Am. Chem. Soc., 127, 8307, 2005.

[3] R. S. Johnson, T. Yamazaki, A. Kovalenko, H. Fenniri, J. Am. Chem. Soc. *129,* 5735, 2007; Editors' choice of Science, 316, 663, 2007.

[4] G. Tikhomirov, T. Yamazaki, A. Kovalenko, H. Fenniri, Langmuir, 24, 4447, 2007.

[5] T. Yamazaki, N. Blinov, D. Wishart, A. Kovalenko, Biophys. J., 95, 4540, 2008.

[6] T. Luchko, S. Gusarov, D. R. Roe, C. Simmerling, D. A. Case, J. Tuszynski, A. Kovalenko, J. Chem. Theory Comput., 6, 607, 2010.

[7] N. Blinov, L. Dorosh, D. Wishart, A. Kovalenko, Biophys. J. 98, 282, 2010.

[8] T. Imai, K. Oda, A. Kovalenko, F. Hirata, A. Kidera, J. Am., Chem. Soc., 131, 12430, 2009.

[9] N. Yoshida, T. Imai, S. Phongphanphanee, A. Kovalenko, F. Hirata, J. Phys. Chem. B Feature Article, 113, 873, 2009.

[10] M. C. Stumpe, N. Blinov, D. Wishart, A. Kovalenko, V. S. Pande, (2010), submitted.

[11] N. Blinov, R. Perez Pineiro, T. Bjorndahl, D. Wishart, A. Kovalenko, to be published.

[12] J. S. Perkyns, B. M. Pettitt, J. Chem. Phys., 97, 7656, 1992.

[13] T. Yamazaki, A. Kovalenko, J. Chem. Theory Comput., 5, 1723, 2009.

[13] J. W. Kaminski, S. Gusarov, T. A. Wesolowski, A. Kovalenko, J. Phys. Chem. B, 2010, (in press).