

The development of a revised version of multi-center molecular Ornstein-Zernike equation

Kentaro Kido^a, Daisuke Yokogawa^b and Hirofumi Sato^{a,*}

^a*Department of Molecular Engineering, Kyoto University, Nishikyo, Kyoto,
615-8510 Japan*

^b*Department of Chemistry, Nagoya University, Furo-cho, Chikusa-ku, Nagoya,
464-8602 Japan*

Abstract

Ornstein-Zernike (OZ)-type theory is a powerful tool to obtain 3-dimensional solvent distribution around solute molecule. Recently, we proposed multi-center molecular OZ method, which is suitable for parallel computing of 3D solvation structure. The distribution function in this method consists of two components, namely reference and residue parts. Several types of the function were examined as the reference part to investigate the numerical robustness of the method. As the benchmark, the method is applied to water, benzene in aqueous solution and single-walled carbon nanotube in chloroform solution. The results indicate that fully-parallelization is achieved by utilizing the newly proposed reference functions.

Key words: OZ-type equation, multi-center molecular OZ (MC-MOZ), SW-CNT

PACS:

1 Introduction

The integral equation theory for molecular liquids [1,2] is a powerful tool to obtain liquid and solvation structures. In the three-dimensional (3D) formalism, solvent distribution in the vicinity of the intended molecule is provided by Ornstein-Zernike (OZ)-type equations. The 3D reference interaction site model (3D-RISM) [3] is one of the representatives of the approach. The accuracy and validity of the method are demonstrated in a variety of applications including simple atomic ion and biological systems like protein in solution phase [4–7]. 3D-RISM is capable of adequately computing the solvation structure not only around a protein but also inside.

Recently, an alternative approach called multi-center molecular OZ (MC-MOZ) method is proposed [8–10]. In MC-MOZ, a solvation structure is expressed as the sum of two terms, namely reference and residue terms. After obtaining the former term by solving the standard RISM equation for radial distribution, the latter residue term representing the angular distribution is computed. Because the residue term is described as a set of distribution functions, which are localized in the vicinity of each atom (site) in the solute molecule such as protein, the total parallel efficiency was in practice lowered as the number of processors was increased.

In the present study, several other functions were examined as the reference term to investigate the numerical robustness of the method. Although one can arbitrarily set the reference term, the computed 3D distribution function should in principle be obtained uniquely. More specifically, the residue term is

* Hirofumi Sato

Email address: `hirofumi@moleng.kyoto-u.ac.jp` (Hirofumi Sato).

expected to complement the difference caused by the reference. Another aim of the present study is to achieve the full-parallelization of the computation. The solvation structures of water (TIP3P-like), benzene in aqueous solution and single-walled carbon nanotube (SW-CNT, 1024 atoms) in chloroform solution are calculated using the new reference, and the obtained 3D solvent distributions are compared with the original MC-MOZ procedure.

2 Method

The detail of MC-MOZ has been described in Ref. [8]. It is our intent to assume the readers' familiarity of MC-MOZ as well as the statistical mechanics of molecular liquids; otherwise, textbooks such as [11–13] and other references [1–3,14] are available. Here, only the reference terms are focused on because the corresponding residue terms are computed through the same procedure in Ref. [8]. On the computations, two types of functions are prepared as the reference term.

$$h_s^{\text{ref}}(\mathbf{r}) = \sum_{\alpha}^{\text{solute}} \sum_{s'}^{\text{solvent}} c_{\alpha s'}^{\text{method}} * (w_{s's}^{\text{V}} + \rho_s h_{s's}^{\text{V}})(r_{\alpha s}) \quad (1)$$

$$c_s^{\text{ref}}(\mathbf{r}) = \sum_{\alpha}^{\text{solute}} c_{\alpha s}^{\text{method}}(r_{\alpha s}) \quad (2)$$

where $h_s^{\text{ref}}(\mathbf{r})$ and $c_s^{\text{ref}}(\mathbf{r})$ are the reference terms of total and direct correlation functions for solvent site s , respectively. $w_{ss'}^{\text{V}}$ and $h_{ss'}^{\text{V}}$ denote the solvent-solvent intramolecular and total correlation functions, respectively. ρ_s is the number density of solvent and the asterisk represents a convolution integral.

In the original MC-MOZ procedure, standard 1-dimensional (1D) RISM is employed to prepare the reference term. $c_{\alpha s}^{\text{1D-RISM}}(r_{\alpha s})$ is obtained by solving

the corresponding solute-solvent 1D-RISM equation [14] coupled with an adequate closure. Then this function is plugged into the right hand side of Eqs. (1) and (2) to evaluate the residue term.

$$c_{\alpha s}^{\text{method}}(r_{\alpha s}) = c_{\alpha s}^{\text{1D-RISM}}(r_{\alpha s}). \quad (3)$$

As mentioned above, the procedure can not be independently performed for each solute site because $c_{\alpha s}^{\text{1D-RISM}}(r_{\alpha s})$ depend on the information of other solute sites. This is the cause for the low parallel efficiency. In the present study, following four functions are examined for $c_{\alpha s}^{\text{method}}$ to prepare the reference term. The first one is just zero;

$$c_{\alpha s}^{\text{I}}(r_{\alpha s}) = 0. \quad (4)$$

$h_s^{\text{ref}}(\mathbf{r})$ is generated using Eq. (1), hence both of $c_s^{\text{ref}}(\mathbf{r})$ and $h_s^{\text{ref}}(\mathbf{r})$ are zero value functions. The second one is based on the analytical solution of OZ/Percus-Yevick (PY) equations for hard spheres [15–17].

$$c_{\alpha s}^{\text{II}}(r_{\alpha s}) = \begin{cases} -\lambda_0 + 6\eta\lambda_2 \left(\frac{r_{\alpha s}}{\sigma_{\alpha s}}\right) - \frac{1}{2}\eta\lambda_1 \left(\frac{r_{\alpha s}}{\sigma_{\alpha s}}\right)^3 & (r_{\alpha s} < \sigma_{\alpha s}) \\ 0 & (r_{\alpha s} \geq \sigma_{\alpha s}) \end{cases} \quad (5)$$

$$\lambda_0 = 6\eta\lambda_2 - \frac{1}{2}\eta\lambda_1, \quad \lambda_1 = \frac{(1+2\eta)}{(1-\eta)^4}, \quad \lambda_2 = \frac{(1+1/2\eta)}{(1-\eta)^4}$$

where $\sigma_{\alpha s} = (\sigma_{\alpha} + \sigma_s)/2$ (σ is the Lennard-Jones (LJ) radius) and η denotes a packing fraction that is set to 0.30 in this study. In a similar manner, $c_s^{\text{ref}}(\mathbf{r})$ and $h_s^{\text{ref}}(\mathbf{r})$ are obtained using Eqs. (1) and (2). The third one is generated by mean spherical approximation (MSA) with the LJ potential

$$c_{\alpha s}^{\text{III}}(r_{\alpha s}) = \begin{cases} 0 & (r_{\alpha s} < \sigma_{\alpha s}) \\ -\frac{1}{k_B T} u^{\text{LJ}}(r_{\alpha s}) & (r_{\alpha s} \geq \sigma_{\alpha s}) \end{cases} \quad (6)$$

where k_B and T represent Boltzmann constant and temperature. $u^{\text{LJ}}(r_{\alpha s})$ denotes the LJ potential between site α and s . Here, it should be emphasized that these new reference terms no longer depends on the information of other solute sites; namely, these computations can be site-independently performed. Furthermore, they can be prepared in non-iterative manner whereas it is necessary to iteratively solve 1D-RISM equation in the original procedure.

The last one is an iterative but independent of other solute sites. A 1D-RISM-like equation is considered.

$$h_{\alpha s}(r_{\alpha s}) = \sum_{\gamma}^{\text{solute}} \sum_{s'}^{\text{solvent}} w_{\alpha\gamma} * c_{\alpha s'}^{\text{IV}} * (w_{s's}^{\text{V}} + \rho_s h_{s's}^{\text{V}})(r_{\alpha s}), \quad (7)$$

where $h_{\alpha s}(r_{\alpha s})$ and $w_{\alpha\gamma}$ represent solute-solvent total and solute-solute intramolecular correlation functions, respectively. By numerically solving this equation coupled with an adequate closure, $c_{\alpha s}^{\text{IV}}(r_{\alpha s})$ is obtained. Then $h_s^{\text{ref}}(\mathbf{r})$ is calculated using Eq. (1). Note that the index of $c_{\alpha s'}^{\text{IV}}$ in this equation is *not* $\gamma s'$, corresponding to the standard 1D-RISM equation. Although there is no physical meaning of this description, it enables us to compute $c_{\alpha s'}^{\text{IV}}$ in a site-independent manner. Being different from above three methods (I, II and III), $c_{\alpha s}^{\text{IV}}(r_{\alpha s})$ implicitly involves the contribution of solute geometry through the matrix $w_{\alpha\gamma}$ and that of solute-solvent interaction.

The potential parameters used in the present study for water (TIP3P[22,23]-like), benzene [24] and SW-CNT [25,26] are summarized in Table 1. In SW-CNT (the CNT index = (16,0)), the C-C distance and the CCC angle are

1.418Å and 120.0 deg, respectively. The diameter and the cylindrical length are 12.52Å and 68.00Å, respectively. For all calculations, the solution temperature is set to 298.15K. The number densities of water and chloroform are 1.000 g/cm³(= 0.033426 molecule/Å³) and 1.479 g/cm³(= 0.007460 molecule/Å³) [35], respectively.

3 Results and discussion

3.1 Water and benzene

Three-dimensional hydration structures around water (Figure 1) and benzene (Figure 2) molecules are obtained by MC-MOZ theory coupled with HNC closure. The degree, l , is set to 14 for the real spherical harmonics. 512 (logarithm) and 302 (Lebedev) grid points are used as the radial and angular grids, respectively.

The final results, total correlation functions, were obtained as the sum of the reference and residue terms ($h_s^{1D-RISM}(\mathbf{r}_k)$, $h_s^I(\mathbf{r}_k)$, $h_s^{II}(\mathbf{r}_k)$, $h_s^{III}(\mathbf{r}_k)$ and $h_s^{IV}(\mathbf{r}_k)$) after the computation. Figure 1 shows the hydration structures near water molecule computed by all the reference terms: the original (panel A), method I (panel B), II (panel C), III (panel D) and IV (panel E). They look quite similar to each other. In order to quantify the deviation, mean absolute deviation (MAD) on all the grid points is introduced,

$$\text{MAD} = \sum_s^{N_v} \sum_k^{N_g} \frac{|h_s^i(\mathbf{r}_k) - h_s^{1D-RISM}(\mathbf{r}_k)|}{N_v N_g} \quad (i = \text{I, II, III and IV}), \quad (8)$$

where N_v and N_g denote the numbers of solvent site and total grid points

(radial part $\times$ angular part), respectively. Note that MAD is the averaged difference of $h_s^i(\mathbf{r})$ with respect to $h_s^{1D-RISM}(\mathbf{r})$ on all over the grid points used in the computation. These values of the method I, II, III and IV are 2.12×10^{-3} , 2.54×10^{-3} , 2.19×10^{-3} and 3.03×10^{-3} , respectively. They are sufficiently small, hence all the new reference terms provide essentially the same result to the original procedure. In fact, the excess chemical potentials of hydration (hydration free energies, HFEs) and partial molar volumes (PMVs) computed by integration of total and direct correlation functions [18–20], are virtually identical; HFEs are -2.29 kcal/mol (original), -2.29 kcal/mol (I), -2.30 kcal/mol (II), -2.29 kcal/mol (III) and -2.29 kcal/mol (IV). And PMVs are 14.6 cm³/mol (original), 14.6 cm³/mol (I), 14.6 cm³/mol (II), 14.6 cm³/mol (III) and 14.6 cm³/mol (IV). The experimental values of HFE and PMV are -6.3 kcal/mol [27,28] and 18.0 cm³/mol [29], respectively. Since molecular simulation study using essentially the same potential parameters provides accurate HFE [30], these errors are probably caused by the HNC closure, which does not accurately describe an exclusive volume of solvent molecule [31–34]. We have checked the effect of the tolerance to solve 1D-RISM equation for judging the convergence and l on MAD. The former (10^{-10} for all calculations in the present study) does not affect MAD at all, and the latter shows slight dependency. For example, the MAD by method I with respect to $h_s^{1D-RISM}$ ($l = 14$) is gradually improved by increasing l : 3.69×10^{-3} ($l = 11$), 3.34×10^{-3} ($l = 12$) and 3.16×10^{-3} ($l = 13$).

Hydration structures near benzene molecule are presented in Figure 2. Being similar to the water molecule case, the structures by new reference terms (panels B, C, D and E) resemble that by the original one (panel A). The MADs are 1.42×10^{-2} (I), 1.78×10^{-2} (II), 1.57×10^{-2} (III) and 6.92×10^{-3} (IV),

respectively. In comparison with the corresponding values of water case, they are slightly greater but hardly affect the HFE estimations; 12.21 (original), 12.21 (I), 12.22 (II), 12.21 (III) and 12.21 kcal/mol (IV). Similarly, PMVs are 54.0 (original), 54.1 (I), 54.0 (II), 54.0 (III) and 54.1 cm³/mol (IV).

These facts show that MC-MOZ is numerically robust enough, at least for these small molecule systems, and any kinds of initial function can be adopted as the reference term. While the new reference function can be prepared non-iteratively, the residue term is determined through a numerical iterative procedure to satisfy the OZ-type equation. The numbers of iterations are different among each solute site, because the residue term assigned to this site is independently computed as mentioned above. In the case of water molecule, the averaged iteration numbers are 254 (original), 281 (I), 280 (II), 210 (III) and 207 (IV). Similarly for the benzene calculation, the numbers are 203 (original), 197 (I), 197 (II), 200 (III) and 200 (IV). These results suggest that the convergence does not strongly depend on the choice of the reference term.

3.2 SW-CNT

Kovalenko-Hirata closure[21] is employed to compute the solvation structure around the SW-CNT molecule. As the degree of a real spherical harmonics $l = 11$ is used. The numbers of radial and angular grid points are 512 and 194, respectively. The SW-CNT calculations were performed by CPUs of SGI Altix in Institute for Molecular Science (Okazaki Japan).

To check the parallel efficiency of the calculation, speed up ratio, $S(n)$, is defined as follows:

$$S(n) = 2 \times \frac{[\text{execution time with 2 processors}]}{[\text{execution time with } n \text{ processors}]}, \quad (9)$$

where n is the number of processors. If parallelization is perfect, $S(n)$ is identical with n . $S(n)$ for the SW-CNT calculations are shown in Figure 3. As expected, the present method (I, open triangles) realizes remarkable parallelizations. On the other hand, $S(n)$ values by the original procedure (open squares) reach a plateau on $n > 16$. As pointed out in Ref. [8], each processor must summarize large-size data from other sites in the each iteration step on the construction of the 1D-RISM reference term (closed diamonds). So this process basically determines the total speed up ratio, especially when a large number of processors is used. In contrast, the procedure of residue term is well-parallelized (open diamonds). As the results, almost all the total execution times (more than 80%) are consumed by the reference term routine, which makes the parallelization less efficient.

Finally, let us discuss the solvation structure around SW-CNT. Different from the cases of water and benzene molecules, the distributions obtained by method I, II and III do not closely resemble that of the original 1D-RISM reference. Their MADs are 1.16×10^{-1} (I), 2.00×10^{-1} (II) and 9.91×10^{-2} (III). These values are indeed larger than those of water and benzene molecules. On the other hand, the MAD of method IV is sufficiently small, 1.10×10^{-2} , which is similar to the benzene molecule case.

The result indicates that starting from the original 1D-RISM and from method IV the same result is reached while the others give different numerical solutions. It is difficult to fully understand this behavior but one possible explanation is the insufficiency of l . In other words, the numerical solution of MC-MOZ procedure could not depend on the reference term in the limit of an

infinite number of l . In order to confirm this, we also performed an additional calculation using a greater $l(= 14)$ and a finer angular grid (the number of points was 302 for a site). Figure 4 shows the distributions of the CH site in chloroform solvent near SW-CNT. MAD of method IV (panel E) becomes sufficiently small (9.27×10^{-3}), indicating that the numerical solution of method IV is virtually equal to the original one (panel A). Similarly, MADs of method I (panel B), II (panel C) and III (panel D) are decreased; 9.43×10^{-2} , 1.59×10^{-1} and 8.24×10^{-2} , respectively. In fact, the improvement is $\sim 20\%$. Therefore, the deviation seems to come from rather slow convergence on l . It should be emphasized that the original MC-MOZ gives the same result to 3D-RISM, as shown in the previous study[8]. By solving 1D-RISM equation, many-body effect is undoubtedly integrated into the reference term. Methods I and III are, however, purely site-independent procedure, meaning that the effect from other sites is not taken into account at all. For simple molecules like water, this insufficiency is recovered by the residue term. In contrast, both of methods II and IV deal with many-body effect (Note that method II utilizes OZ/PY analytical solution for hard sphere system). A possible reason for the superiority of method IV may come from the intramolecular correlation function both for solute-solute and solvent-solvent. The other reason is the effect of partial charges on solvent sites.

4 Conclusion

The four functions are examined as the reference part to investigate the numerical robustness of MC-MOZ method. As benchmark tests, the hydration structures around water and benzene molecules are computed. All the refer-

ence functions give virtually the same distribution and thermodynamic quantities. Thanks to the site-independent character of the newly proposed references, fully-parallelization is achieved by utilizing them. On the computation of the solvation structure around SW-CNT, the speed up ratio of the proposed procedure shows a linear scaling for a number of processors. A comparison of the solvation structures near SW-CNT suggests that the reference term by method IV is the most promising to efficiently compute 3D solvation structure.

Acknowledgments

The work is financially supported in part by Grant-in-Aid for Scientific Research on Priority Areas “Molecular Science for Supra Functional Systems” (477-22018016), Grant-in-Aid for Scientific Research on Innovative Areas “Molecular Science of Fluctuations” (2006-21107511), as well as by Grant-in-Aid for Scientific Research (C) (20550013). The Strategic Programs for Innovative Research (SPIRE), the Computational Materials Science Initiative (CMSI), and the Ministry of Education, Culture, Sports, Science and Technology (MEXT) Japan are also acknowledged.

Theoretical computations were partly performed using Research Center for Computational Science, Okazaki, Japan.

References

- [1] D. Chandler, H.C. Andersen, J. Chem. Phys. 57 (1972) 1930.

- [2] F. Hirata, P.J. Rossky, Chem. Phys. Lett. 83 (1981) 329.
- [3] A. Kovalenko, F. Hirata, Chem. Phys. Lett. 290 (1998) 237.
- [4] A. Kovalenko, F. Hirata, J. Chem. Phys. 112 (2000) 10391.
- [5] A. Kovalenko, F. Hirata, J. Chem. Phys. 112 (2000) 10403.
- [6] N. Yoshida, P. Saree, Y. Maruyama, T. Imai, F. Hirata, J. Am. Chem. Soc. 128 (2006) 12042.
- [7] Y. Kiyota, R. Hiraoka, N. Yoshida, Y. Maruyama, T. Imai, F. Hirata, J. Am. Chem. Soc. 131 (2009) 3852.
- [8] D. Yokogawa, H. Sato, T. Imai, S. Sakaki, J. Chem. Phys. 130 (2009) 064111.
- [9] D. Yokogawa, H. Sato, S. Sakaki, J. Mol. Liq. 147 (2009) 112.
- [10] K. Hirano, D. Yokogawa, H. Sato, S. Sakaki, J. Phys. Chem. B 114 (2010) 7935.
- [11] J.P. Hansen and I.R. McDonald Theory of Simple Liquids (1990) Academic Press.
- [12] F. Hirata (Ed.), Molecular Theory of Solvation (Understanding Chemical Reactivity), Kluwer-Springer, 2004.
- [13] B. Mennucci and R. Cammi (Eds.) Continuum Solvation Models in Chemical Physics: From Theory to Applications, John Wiley & Sons Inc. 2008.
- [14] F. Hirata, P.J. Rossky, B.M. Pettitt, J. Chem. Phys. 78 (1983) 4133.
- [15] M.S. Wertheim, Phys. Rev. Lett. 10 (1963) 321.
- [16] M.S. Wertheim, J. Math. Phys. 5 (1964) 643.
- [17] J.L. Lebowitz, Phys. Rev. 133 (1964) A895.
- [18] S.J. Singer, D. Chandler, Mol. Phys. 55 (1985) 621.

- [19] T. Imai, M. Kinoshita, F. Hirata, J. Chem. Phys. 112 (2000) 9469.
- [20] Y. Harano, T. Imai, A. Kovalenko, M. Kinoshita, F. Hirata, J. Chem. Phys. 114 (2001) 9506.
- [21] A. Kovalenko, F. Hirata, J. Chem. Phys. 110 (1999) 10095.
- [22] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, J. Chem. Phys. 79 (1983) 926.
- [23] W.L. Jorgensen, J.D. Madura, Mol. Phys. 99 (1985) 1381.
- [24] W.L. Jorgensen, J.M. Briggs, M.L. Contreras, J. Phys. Chem. 94 (1990) 1683.
- [25] J.H. Walther, R.L. Jaffe, T. Halicioglu, P. Koumoutsakos, J. Phys. Chem. B 105 (2001) 9980.
- [26] T. Werder, J.H. Walther, R.L. Jaffe, T. Halicioglu, P. Koumoutsakos, J. Phys. Chem. B 107 (2003) 1345.
- [27] C.H. Reynolds, J. Chem. Inf. Comput. Sci. 35 (1995) 738.
- [28] C.J. Cramer, D.G. Truhlar, J. Comput.-Aided Mol. Des. 6 (1993) 629.
- [29] D.P. Kharakoz, J. Sol. Chem. 21 (1992) 569.
- [30] W.L. Jorgensen, J. Tirado-Rives, Perspect. Drug Discov. 3 (1995) 123.
- [31] D. Chandler, Mol. Phys. 31 (1976) 1213.
- [32] P.T. Cummings, C.G. Gray, D.E. Sullivan, J. Phys. A 14 (1981) 1483.
- [33] M. Kinoshita, F. Hirata, J. Chem. Phys. 106 (1997) 5202.
- [34] A. Kovalenko, F. Hirata, J. Chem. Phys. 113 (2000) 2793.
- [35] D.R. Lide (Ed.), CRC Handbook of Chemistry and Physics, 87th ed. CRC press: Boca Raton, FL, (2003) p8-540.

Table 1

Potential parameters for water(TIP3P-like), chloroform, benzene and SW-SNT.

molecule	site	charge / e	σ / Å	ϵ / kcal mol ⁻¹
water (TIP3P[22,23]-like)	Ow	-0.8340	3.151	0.1520
	Hw	0.4170	1.000	0.0560
benzene[24]	C	-0.1030	3.550	0.0700
	H	0.1030	2.420	0.0300
SW-CNT	C	0.0000	3.851	0.1050
chloroform[24]	CH	0.4200	3.800	0.0800
	Cl	-0.1400	3.470	0.3000

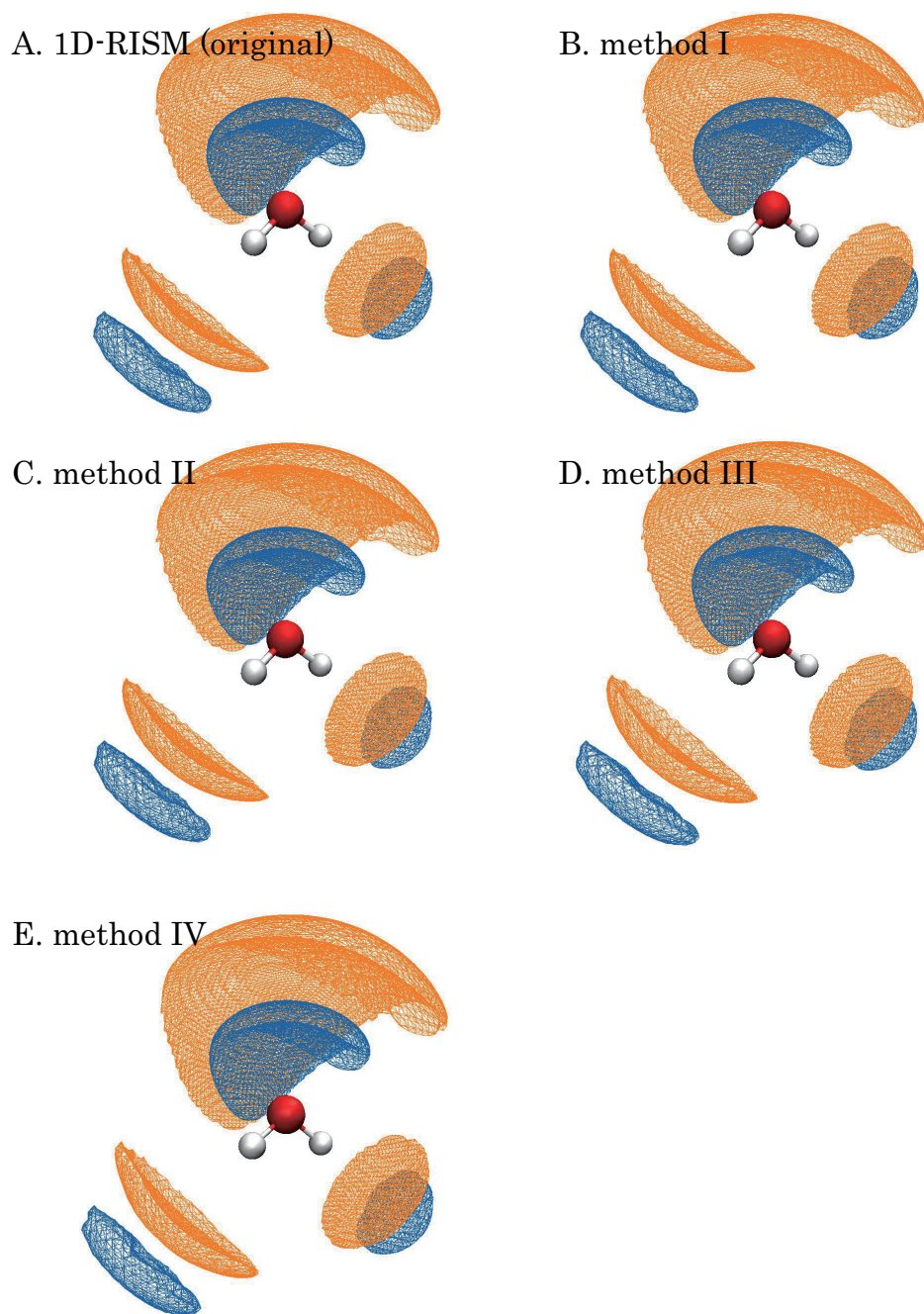
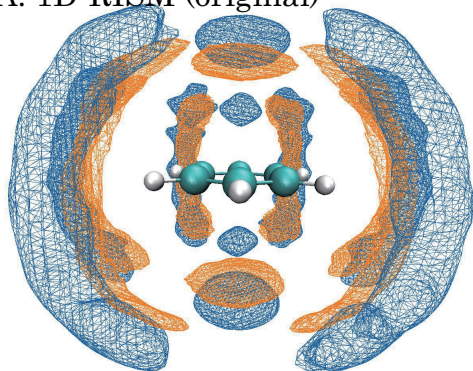
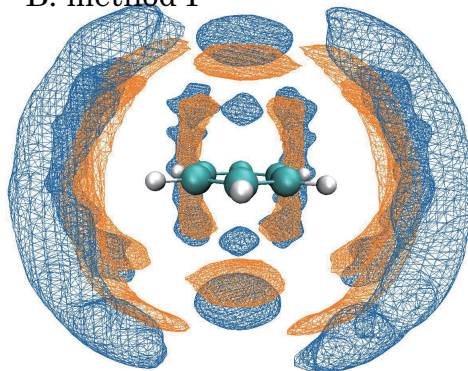


Figure 1. Three-dimensional hydration structures near water molecule computed by original (1D-RISM, A), method I (B), II (C), III (D) and IV (E) reference terms. Blue (isovalue: 1.8) and orange (isovalue: 3.0) iso-surfaces represent the distributions of hydrogen and oxygen atoms of solvent water molecules, respectively.

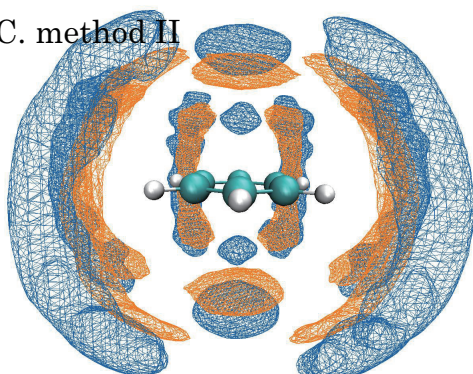
A. 1D-RISM (original)



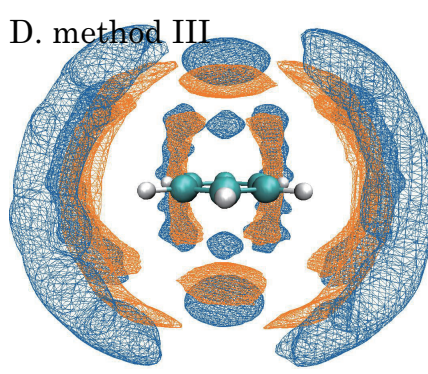
B. method I



C. method II



D. method III



E. method IV

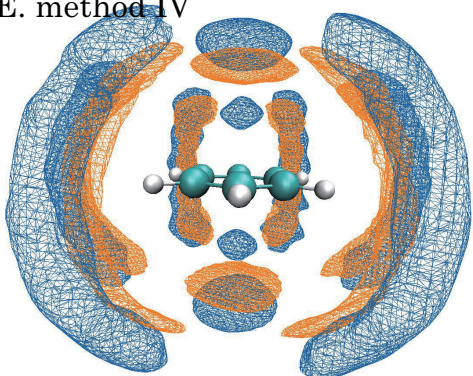


Figure 2. Three-dimensional hydration structures near benzene molecule computed by original (1D-RISM, A), method I (B), II (C), III (D) and IV (E) reference terms. Blue (isovalue: 1.5) and orange (isovalue: 3.0) iso-surfaces represent the distributions of hydrogen and oxygen atoms of solvent water molecules, respectively.

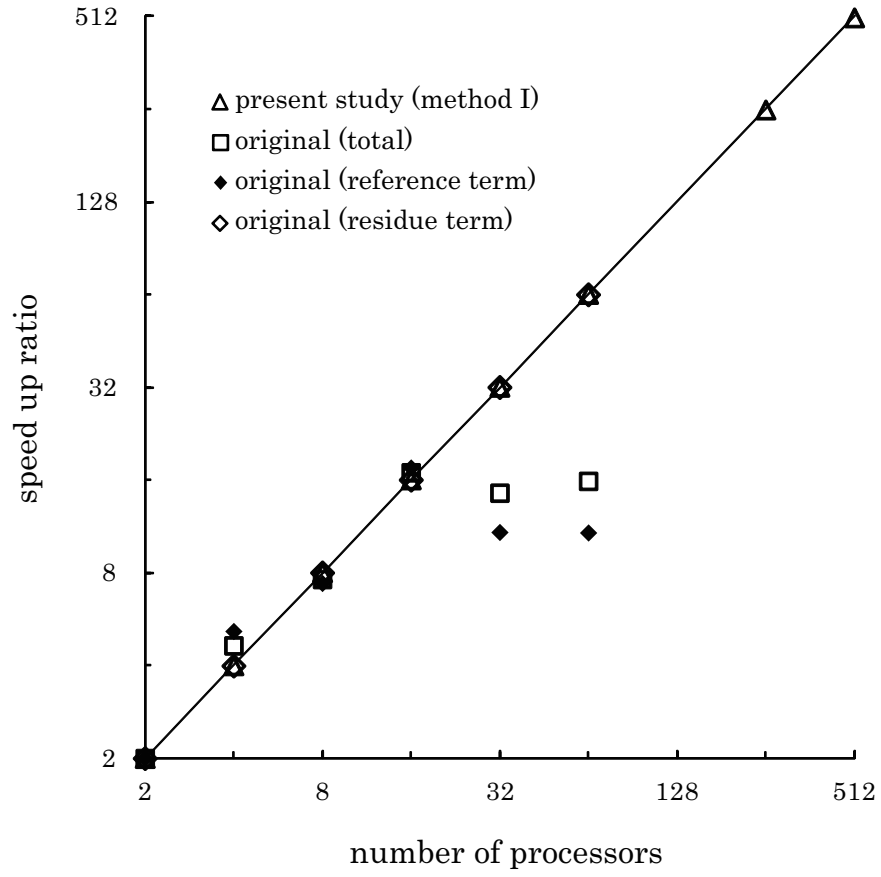


Figure 3. Parallel efficiencies in SW-CNT calculations using the original (open squares) and method I (open triangles) as reference. The subcomponents of the original procedure (reference and residue) are shown.

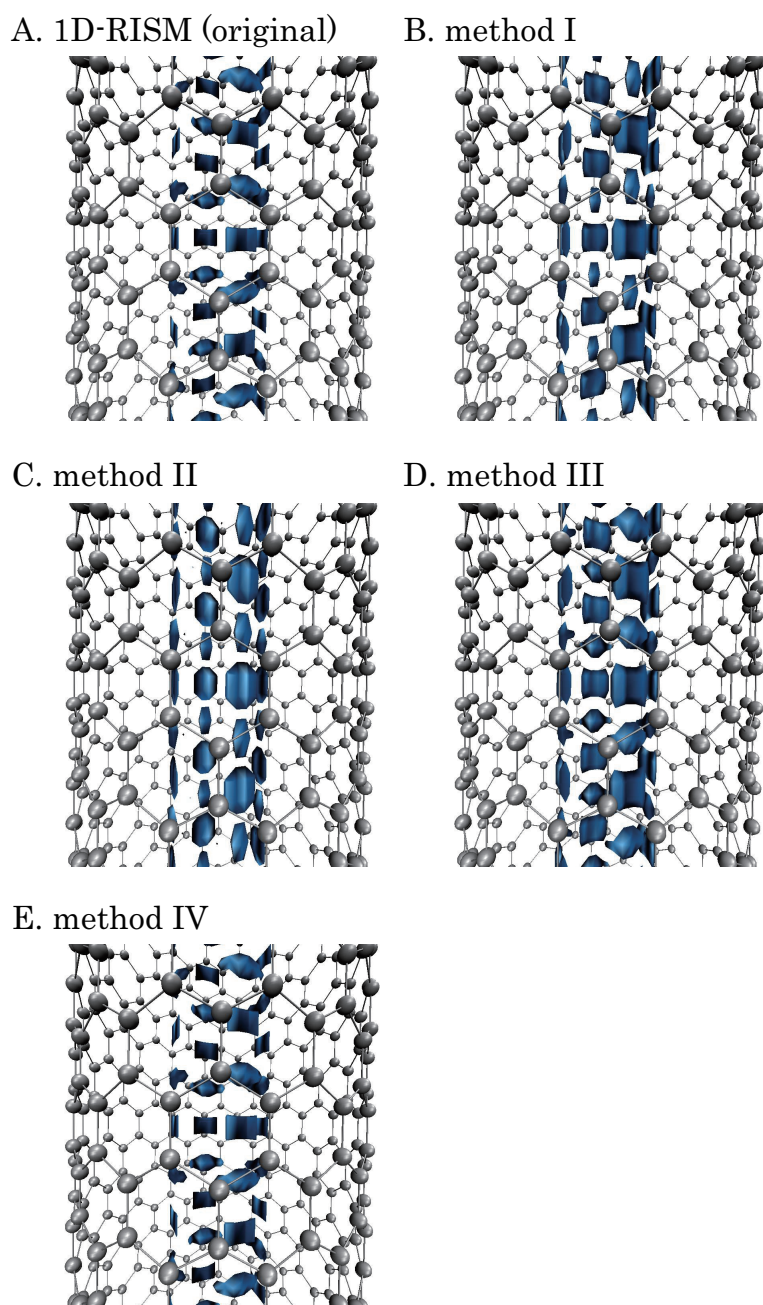


Figure 4. Three-dimensional solvation structures inside SW-CNT calculated by original (panel A), method I (panel B), II (panel C), III (panel D) and IV (panel E) reference terms. Iso-surfaces (blue, isovalue: 5.2) represent the distributions of CH site of solvent chloroform molecule.