

# A new method to reconstruct three-dimensional spatial distribution function from radial distribution function in solvation structure

Daisuke Yokogawa, Hirofumi Sato,<sup>a)</sup> and Shigeyoshi Sakaki

Department of Molecular Engineering, Graduate School of Engineering, Kyoto University, Nishikyo-ku, Kyoto 615-8510, Japan

(Received 19 August 2005; accepted 14 October 2005; published online 6 December 2005)

Three-dimensional spatial distribution function (SDF) of solvent is a fundamental quantity for analysis of solvation. However, its calculation has been very limited because long computational time is required. We here developed a novel and robust method to construct approximated SDFs of solvent sites from radial distribution functions. In this method, the expansion of SDFs in real solid harmonics around atoms of solute leads to a linear equation, from which SDFs are evaluated with reasonable computational time. This method is applied to the analysis of the solvation structure of liquid water, as an example. The successful results clearly show that this method is very powerful to investigate solvation structure. © 2005 American Institute of Physics. [DOI: 10.1063/1.2137695]

## I. INTRODUCTION

To understand most of the chemical reactions correctly, solvent effects should be considered in theoretical analysis. Various methods such as dielectric continuum model, molecular dynamics (MD), and integral equation theory<sup>1</sup> (IET) of liquids have been proposed to investigate solvent effects. These methods are useful to calculate solvation energy. However, the three-dimensional (3D) information on solvent coordination such as a spatial distribution function (SDF) have not been studied thoroughly, except for limited pioneering works,<sup>2–10</sup> despite the 3D information being very helpful to understand chemical reactions in solution phase by visualizing the solvation feature.<sup>11</sup>

One of those works was reported by Soper *et al.*<sup>7–9</sup> They expanded the SDF as a function of position vector and orientation of solvent using spherical harmonic functions, and optimized the coefficients, which determine the shape of SDF, with the minimum noise formalism.<sup>9</sup> The equation for the coefficients was solved in an iterative manner. Sato *et al.*<sup>3</sup> presented the “most plausible solvation structure” (MPSS) using the radial distribution functions (RDFs). Simulation techniques such as the MD method, and three-dimensional reference interaction model (3D-RISM),<sup>12,13</sup> can evaluate SDF directly. However, both of the methods need long computational time to calculate SDF. There are other IETs, such as MOZ theory,<sup>14–21</sup> that can analyze three-dimensional structure of solvation, though those results have the approximations inherent in IETs.

On the other hand, RDFs, which are the most frequently used in discussion, are easily evaluated with much shorter computational time. Thus, it is highly desired to develop the method that easily provides SDF from RDFs.

In this communication, we newly propose an interesting method to reconstruct approximated SDFs of solvent site from RDFs, which are calculated by any solvation theory, by employing the spherical harmonic expansion around each

solute site. This expansion leads to simple linear equation and we can obtain the coefficients determining the shape of approximated SDFs by solving the equation. The efficiency of this method is clearly shown here by applying this method to the coordination of solvent water around a water molecule.

## II. METHOD AND COMPUTATIONAL DETAILS

### A. Method

We begin with SDFs of solvent site  $s$  around a solute molecule,  $n_s(\mathbf{r})$ ,

$$\rho_s(\mathbf{r}) = \rho n_s(\mathbf{r}), \quad (1)$$

where  $\rho$  is the number density of solvent and  $\mathbf{r}$  is the position vector in 3D space. The RDF between solute site  $\eta$  and solvent site  $s$  is related to SDFs by Eq. (2),

$$g_{\eta,s}(R) = \frac{1}{4\pi R^2} \int n_s(\mathbf{x} + \mathbf{Q}_\eta) \delta(|\mathbf{x}| - R) d\mathbf{x}, \quad (2)$$

where  $\mathbf{Q}_\eta$  is the position vector of the  $\eta$  site and  $R$  represents the distance between the  $\eta$  site and the  $s$  site.

The SDF of solvent  $s$  site is well approximated by basis functions including real solid harmonics  $S_{lm}$  of the center which is located on individual solute site  $\eta$ ,

$$n_s^\eta(\mathbf{r}) = \left[ \sum_i^N f_i^{\eta,s}(|\mathbf{r} - \mathbf{Q}_\eta|) \sum_{l,m} C_{l,m}^{\eta,s,i} S_{l,m} \left( \frac{\mathbf{r} - \mathbf{Q}_\eta}{|\mathbf{r} - \mathbf{Q}_\eta|} \right) \right] + f_f^{\eta,s}(|\mathbf{r} - \mathbf{Q}_\eta|), \quad (3)$$

where  $\{C_{l,m}^{\eta,s,i}\}$  are the coefficients and  $f_i^{\eta,s}$  and  $f_f^{\eta,s}$  are the basis functions in radial part up to  $N$ . After substitution of Eq. (3) into Eq. (2) and integration over the angular part, we get

<sup>a)</sup>Electronic mail: hirofumi@moleng.kyoto-u.ac.jp

$$g_{\eta,s}(R) = \sum_i^N f_i^{\eta,s}(R) C_{0,0}^{\eta,s,i} + f_f^{\eta,s}(R), \quad (4)$$

where  $\{C_{0,0}^{\eta,s,i}\}$  are fixed as 1.<sup>22</sup> The basis functions  $f_i^{\eta,s}(R)$  and  $f_f^{\eta,s}(R)$  in Eqs. (4) are determined to reproduce  $g_{\eta,s}(R)$ .  $f_i^{\eta,s}(R)$  and  $f_f^{\eta,s}(R)$  are used to represent solvation shell and bulk solvent region, respectively. The function  $f_f^{\eta,s}(R)$  converges to 1 in  $R \rightarrow \infty$ . Hereafter we call the RDF calculated by solvation theory as reference RDF (RRDF). If we can expand RRDFs exactly by  $f_i^{\eta,s}(R)$  and  $f_f^{\eta,s}(R)$ , RRDFs are reproduced correctly by integrating over the angular part of  $n_s^\eta(\mathbf{r})$  in Eq. (3) around solute  $\eta$  site.

It is noted that in each SDF for solvent  $s$  site, there are expanded SDFs to the number of solute site  $N_a$ . The coefficients  $\{C_{l,m}^{\eta,s,i}\}$  in Eq. (3) are determined so that the difference between all these SDFs centered on different solute sites becomes the smallest; when a sufficiently large number of real solid harmonics and basis functions are employed, all of the SDFs become almost unique. Here, we define the error function  $\Gamma_s$  as follows:

$$\Gamma_s = \frac{1}{2} \sum_k^{N_p} \sum_{\eta \neq \gamma}^{N_a} (n_s^\eta(\mathbf{r}_k) - n_s^\gamma(\mathbf{r}_k))^2, \quad (5)$$

where  $N_p$  is the number of the grid points which are prepared around the solute. As the number of basis functions in Eq. (3) increases,  $\Gamma_s$  decreases. By minimizing Eq. (5), we can obtain the equation that determines the coefficients  $\{C_{l,m}^{\eta,s,i}\}$ ,

$$\mathbf{C} = -\mathbf{A}^{-1}\mathbf{B}, \quad (6)$$

where  $\mathbf{C}$  is the vector whose component is  $C_{l,m}^{\eta,s,i}$  and the components of  $\mathbf{A}$  and  $\mathbf{B}$  are given by Eqs. (7a) and (7b),

$$A_{\eta,j,l',m'}^{\gamma,s,i,l,m} = (N_a \delta_{\eta,\gamma} - 1) \sum_{k=1}^{N_p} f_i^{\eta,s}(|\mathbf{r}_k - \mathbf{Q}_\eta|) f_j^{\gamma,s}(|\mathbf{r}_k - \mathbf{Q}_\gamma|) \times S_{l,m} \left( \frac{\mathbf{r}_k - \mathbf{Q}_\eta}{|\mathbf{r}_k - \mathbf{Q}_\eta|} \right) S_{l',m'} \left( \frac{\mathbf{r}_k - \mathbf{Q}_\gamma}{|\mathbf{r}_k - \mathbf{Q}_\gamma|} \right), \quad (7a)$$

$$B^{\eta,s,i,l,m} = \sum_{k=1}^{N_p} f_i^{\eta,s}(|\mathbf{r}_k - \mathbf{Q}_\eta|) S_{l,m} \left( \frac{\mathbf{r}_k - \mathbf{Q}_\eta}{|\mathbf{r}_k - \mathbf{Q}_\eta|} \right) \times \sum_{\gamma} (N_a \delta_{\eta,\gamma} - 1) f_f^{\gamma,s}(|\mathbf{r}_k - \mathbf{Q}_\gamma|). \quad (7b)$$

By solving this linear equation, we can obtain the coefficients without any iterative calculation. In this regard, the present method provides approximated SDFs very easily.

Because we cannot use infinite number of basis functions in practice, the quality of  $n_s^\eta(\mathbf{r})$ s is not uniform in all space; in other words,  $n_s^\eta(\mathbf{r})$  is given in high accuracy where  $\mathbf{r}$  is close to solute site  $\eta$ , but it gets worse when  $\mathbf{r}$  is distant from  $\eta$  site. To obtain well-balanced SDF  $n_s(\mathbf{r})$  from SDFs in Eq. (3), we used arithmetic average of  $n_s^\eta(\mathbf{r})$ ,

$$n_s(\mathbf{r}) = \sum_{\alpha}^{N_a} n_s^\alpha(\mathbf{r}) / N_a. \quad (8)$$

The computational procedure is summarized as follows: (1) the first step is to fit  $f_{if}^{\eta,s}(R)$  to RRDFs, (2) the second step

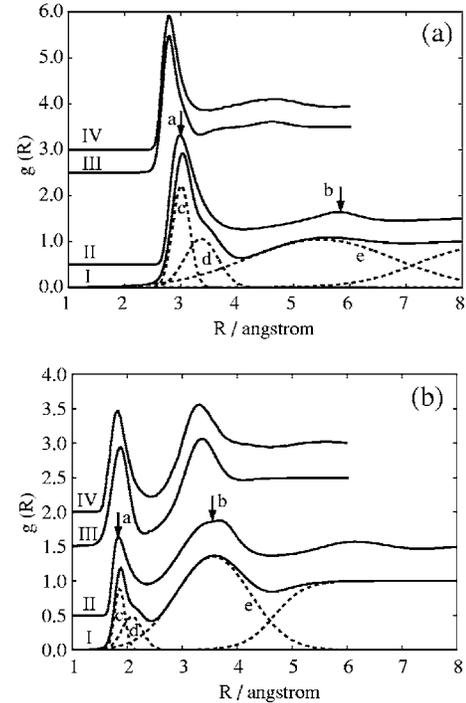


FIG. 1. Fitted RDFs and RRDFs: (a) oxygen-oxygen and (b) hydrogen-oxygen. Solid lines I, II, III, and IV represent fitted RDF(XRISM), RRDF(XRISM), fitted RDF(MD), and RRDF(MD), respectively. The dashed lines represent basis functions used for the fit of RRDF(XRISM).

is to evaluate Eqs. (7a) and (7b) and to solve Eqs. (6) and (3) the third is to take average  $n_s^\eta(\mathbf{r})$  with Eq. (8).

## B. Computational details

We employed the RRDFs calculated by the MD method reported by Jorgensen *et al.*<sup>23</sup> Also, we evaluated here RRDFs with the extended RISM (XRISM),<sup>24</sup> where the simple point charge (SPC)-like water model<sup>25</sup> was employed. The standard modification for the Lennard-Jones (LJ) parameters was made in the hydrogen site. The hypernetted chain (HNC)-like closure was used in solving the XRISM equation. All calculations were carried out at the temperature of 298.15 K and the number density of 0.033 426 molecule/ $\text{\AA}^3$ . Hereafter, we call the RRDFs of MD and XRISM as RRDFs(MD) and RRDFs(XRISM), respectively.

We used Gaussian functions for  $f_i^{\eta,s}(R)$  and the following function of Eq. (9) for  $f_f^{\eta,s}(R)$ :

$$f_f^{\eta,s}(R) = \frac{1}{\exp(-\alpha^{\eta,s}(R - \beta^{\eta,s})) + 1}, \quad (9)$$

to approximate RRDFs in the least-square fitting technique. In this calculation, we employed the real solid harmonics up to  $l=10$ .

## III. RESULTS AND DISCUSSION

The RRDFs (XRISM) and RRDFs(MD) of oxygen-oxygen and hydrogen-oxygen are shown in Fig. 1(a) as lines II and IV and Fig. 1(b) as lines II and IV, respectively. The RDFs of the XRISM (I) and the MD (III) methods were constructed by the fitted basis functions, as shown in Figs.

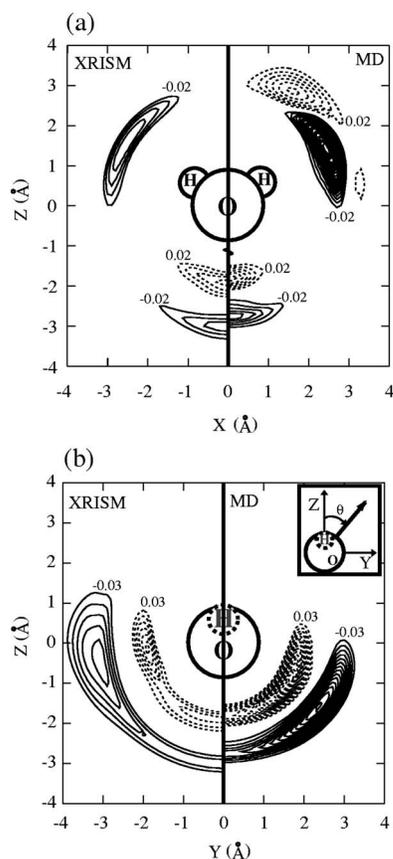


FIG. 2. Charge distribution ( $|e|/\text{\AA}^3$ ): (a) in the plane including the oxygen and two hydrogens and (b) in the bisector plane of HOH angle. The solid line represents the negative part (oxygen), and dashed line is for the positive part (hydrogen). The charge distributions constructed by MD and XRISM are shown in the region  $X > 0$  and the region  $X < 0$ , respectively.

1(a) and 1(b), in comparison with II and IV. We found that three  $f_i^{\eta,s}$  functions and one  $f_f^{\eta,s}$  function are enough to reproduce well RRDFs (II and IV).

However, we cannot reproduce the second peak [b in Figs. 1(a) and 1(b)] properly, when using only two  $f_i^{\eta,s}$  functions and one  $f_f^{\eta,s}$  function. One can see that I and III well reproduce II and IV, respectively. The basis sets c and e were mainly used to reproduce the peaks a and b in RRDFs. The second basis set d seems to correspond to the distribution of the “interstitial water molecules,” which is suggested to be around  $R=3.5 \text{ \AA}$ .<sup>5</sup>

The resultant solvent coordinations in 3D,  $n_O(\mathbf{r})$  and  $n_H(\mathbf{r})$ , are used to calculate the charge density by the following equation:

$$\rho_q(\mathbf{r}) = \rho q_O n_O(\mathbf{r}) + 2\rho q_H n_H(\mathbf{r}), \quad (10)$$

where  $q_O$  and  $q_H$  are the charges of the oxygen and the hydrogen of solvent water, respectively. The charge density in the plane including all the atoms (XZ plane) is shown in Fig. 2(a) and that in the bisector plane of the HOH angle (YZ plane) is shown in Fig. 2(b). Because these density maps have  $C_{2v}$  symmetry, we divided them into two regions and show only a half of them in these figures. The right-hand side is for the MD method and the left-hand side for the XRISM. The negative density (solid line) represents  $n_O(\mathbf{r})$  and the positive density (dashed line) represents  $n_H(\mathbf{r})$ . In Fig. 2(a),

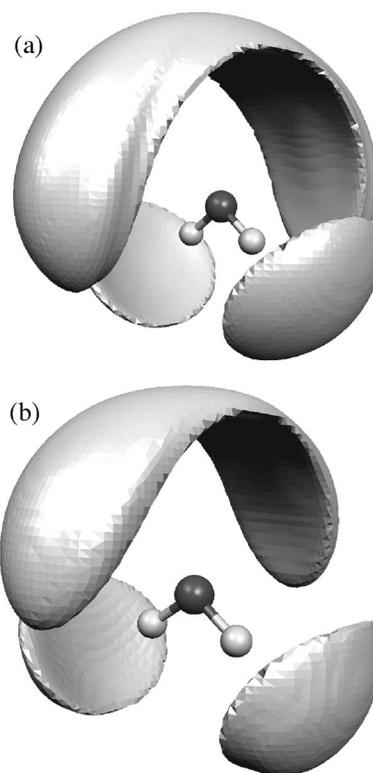


FIG. 3. Three-dimensional SDF map of the oxygen in solvent water using the real solid harmonics up to  $l=10$ : (a) constructed from RRDF(XRISM) and (b) constructed from RRDF(MD). The isodensity surface for  $n_O(\mathbf{r}) = 1.8$  is shown. This figure was drawn with the help of MOLEKEL package (Ref. 29).

one can see the negative charge distribution around the hydrogen of the central water (hereafter, we call it “solute water”). When RRDFs(XRISM) were employed, the maximum of the distribution is located at  $X=-2.5 \text{ \AA}$ ,  $Z=1.6 \text{ \AA}$  and the distance between the maximum and the hydrogen of solute water is evaluated to be  $2.0 \text{ \AA}$ . When RRDFs(MD) were employed, the maximum is at  $X=2.4 \text{ \AA}$ ,  $Z=1.4 \text{ \AA}$  and the distance is  $1.8 \text{ \AA}$ . Thus, these distributions are attributed to the first peak in Fig. 1(b). These negative charge distributions arise from the oxygen that forms hydrogen bond with solute water. Positive charge distribution and negative one are at  $X=0.0$ ,  $Z=-2.0 \text{ \AA}$  and  $X=0.0$ ,  $Z=-3.1 \text{ \AA}$  when RRDFs(XRISM) are employed and at  $X=0.0$ ,  $Z=-1.8 \text{ \AA}$  and  $X=0.0$ ,  $Z=-2.8 \text{ \AA}$  when RRDFs(MD) are employed. These distributions are attributed to different types of hydrogen bonding solvent, as will be discussed below. Figure 2(b) shows charge distribution on the bisector plane of HOH. The positive distribution and negative one indicate that solvent water molecules are present over the average at the region of  $Z < 0$ . The angle  $\theta$  (defined in the right-upper box) of O distribution is evaluated to be about  $102^\circ$  when RRDFs(XRISM) are employed and  $124^\circ$  when RRDFs(MD) are employed, respectively. The latter value agrees well with previously reported value (about  $130^\circ$ ).<sup>10,14</sup> The SDF of solvent oxygen site reconstructed by RRDFs(MD) present the  $\theta$  value close to  $125.3^\circ$  ( $=\cos^{-1}(-1/\sqrt{3})$ ) which is the exact value when the oxygen takes perfect tetrahedral network. However, the  $\theta$  value calculated with RRDFs(XRISM) is somewhat smaller than that of  $125.3^\circ$ . This is because the

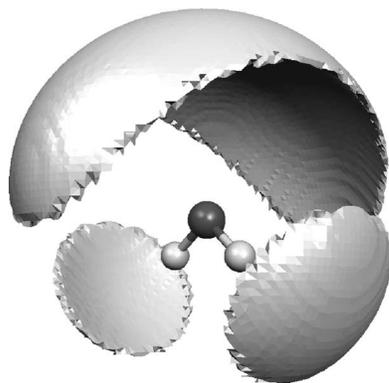


FIG. 4. The three-dimensional SDF map of the oxygen in solvent water constructed by RRDF(XRISM) using the real solid harmonics up to 2. The isodensity surface and the package used are the same as in Fig. 3.

first peak of oxygen shifts to outer region and the first peak of hydrogen shifts to inner region in RRDF(XRISM). We can conclude that this small  $\theta$  value is attributed not to the present reconstruction method but to RRDFs(XRISM). As shown in Figs. 2(a) and 2(b), broad distribution is observed around  $(X, Y, Z) = (0.0, 0.0, -1.9)$ ,  $(0.0, 0.0, -3.0)$  when RRDFs(XRISM) were employed and around  $(X, Y, Z) = (0.0, 0.0, -1.8)$ ,  $(0.0, 0.0, -2.8)$  when RRDFs(MD) were employed. If water coordination took a simple tetrahedron structure (“standard coordination”), there should be no distribution there and two separated distribution could be found at about  $\theta = \pm 125$ . These broad distributions indicate a great deal of variation of solvent water from standard coordination. This continuous distribution of solvent water was also reported by Soper *et al.*<sup>8</sup>

The three-dimensional SDF of  $n_{\text{O}}(\mathbf{r}) > 1.80$  reconstructed by RRDFs(XRISM) and RRDFs(MD) are illustrated in Figs. 3(a) and 3(b), respectively. Both SDFs look very similar to each other. All of the lobes correspond to the distribution of the solvent water that forms hydrogen bond with solute water, as discussed above. This character of the oxygen distribution is essentially the same as those reported by Svishchev *et al.*<sup>10</sup> and Soper *et al.*<sup>8</sup> These results indicate that RRDFs(XRISM) as well as RRDFs(MD) are useful to construct SDFs.

In order to check how much the SDFs depend on real solid harmonic expansion, we evaluated the SDF of solvent oxygen site from RRDF(XRISM) using the real solid harmonics up to  $l=2$ . The reconstructed 3D SDF  $n_{\text{O}}(\mathbf{r})$  is shown as an example in Fig. 4. Although the edge of the lobe is ambiguous because of the insufficient azimuthal accuracy, the shape and the position of the lobes resemble well those of Fig. 3 calculated up to  $l=10$ . The XRISM calculation of water followed by reconstruction of SDF with  $l=2$  is performed in a few minutes in a personal computer.<sup>26</sup> Therefore, the present method to reconstruct SDFs from RDFs is very powerful for investigation of solvation structures.

#### IV. CONCLUSION

A new method to reconstruct SDFs from RDFs is presented here. This method was successfully applied to the liquid structure of water. In this method, the different expansion

from that by Soper *et al.* leads to the linearized equation with which we can easily obtain approximated SDFs. Our method presents reliable results using a small number of real solid harmonics. This means that the present method can be easily applied to large molecular system.

RISM self-consistent field (SCF) can evaluate the solvent structure such as RDFs even for chemical reactions in reasonable computational time.<sup>27,28</sup> The combination of the present method with the RISM-SCF is one of the powerful methods to evaluate the 3D picture of solvation structure.

We will compare this approximated SDFs with SDFs that are directly calculated by MD method in a forthcoming full article.

#### ACKNOWLEDGMENTS

This work has been supported by the Grant-in Aid for Scientific Research on Priority Areas “Water and biomolecules” (430-16041223) from the Ministry of Education, Science, Sports, and Culture (MONBU SHO) in Japan. One of the authors (H.S.) thanks Grant-in Aid for Encouragement of Young Scientists (17750012), the Japanese Ministry of Education, Science, Sports, and Culture (MONBU SHO).

- <sup>1</sup>J. P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, 2nd ed. (Academic, London, 1986).
- <sup>2</sup>C. Nieto-Draghi, J. B. Avalos, and B. Rousseau, *J. Chem. Phys.* **118**, 7954 (2003).
- <sup>3</sup>H. Sato and F. Hirata, *Bull. Chem. Soc. Jpn.* **74**, 1831 (2001).
- <sup>4</sup>I. M. Svishchev and A. Yu. Zassetsky, *J. Chem. Phys.* **113**, 7432 (2000).
- <sup>5</sup>H. Sato and F. Hirata, *J. Chem. Phys.* **111**, 8545 (1999), and references therein.
- <sup>6</sup>A. De Santis and D. Rocca, *J. Chem. Phys.* **107**, 9559 (1997).
- <sup>7</sup>A. K. Soper, *Physica B* **213–214**, 448 (1995).
- <sup>8</sup>A. K. Soper, *J. Chem. Phys.* **101**, 6888 (1994).
- <sup>9</sup>A. K. Soper, C. Andreani, and M. Nardone, *Phys. Rev. E* **47**, 2598 (1993).
- <sup>10</sup>I. M. Svishchev and P. G. Kusalik, *J. Chem. Phys.* **99**, 3049 (1993).
- <sup>11</sup>B. Ensing, F. Buda, M. C. M. Gribnau, and E. J. Baerends, *J. Am. Chem. Soc.* **126**, 4355 (2004).
- <sup>12</sup>M. Ikeguchi and J. Doi, *J. Chem. Phys.* **103**, 5011 (1995).
- <sup>13</sup>D. Beglov and B. Roux, *J. Phys. Chem. B* **101**, 7821 (1997).
- <sup>14</sup>M. Lombardero, C. Martín, S. Jorge, F. Lado, and E. Lomba, *J. Chem. Phys.* **110**, 1148 (1999).
- <sup>15</sup>J. Richardi, C. Millot, and P. H. Fries, *J. Chem. Phys.* **110**, 1138 (1999).
- <sup>16</sup>J. Richardi, P. H. Fries, R. Fischer, S. Rast, and H. Krienke, *Mol. Phys.* **93**, 925 (1998).
- <sup>17</sup>M. Alvarez, F. Lado, E. Lomba, M. Lombardero, and C. Martín, *J. Chem. Phys.* **107**, 4642 (1997).
- <sup>18</sup>P. H. Fries, J. Richardi, and H. Krienke, *Mol. Phys.* **90**, 841 (1997).
- <sup>19</sup>F. Lado, E. Lomba, and M. Lombardero, *J. Chem. Phys.* **103**, 481 (1995).
- <sup>20</sup>P. H. Fries, W. Kunz, P. Calmettes, and P. Turq, *J. Chem. Phys.* **101**, 554 (1994).
- <sup>21</sup>P. H. Fries and G. N. Patey, *J. Chem. Phys.* **82**, 429 (1985).
- <sup>22</sup> $\{C_{0,0}^{n,s,i}\}$  are angular scaling factors of  $f_i^{n,s}$ . The “solvent shell” of RDF is represented by  $f_i^{n,s} \times C_{0,0}^{n,s,i}$ . This means that  $C_{0,0}^{n,s,i}$  value is arbitrarily taken and we fixed  $\{C_{0,0}^{n,s,i}\}$  as 1 in this work.
- <sup>23</sup>W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.* **79**, 926 (1983).
- <sup>24</sup>F. Hirata, B. M. Pettitt, and P. J. Rossky, *J. Chem. Phys.* **77**, 509 (1982).
- <sup>25</sup>H. J. C. Berendsen, J. P. M. Postma, W. F. von Gunstern, and J. Hermas, in *Intermolecular Forces*, edited by B. Pullman (Reidel, Dordrecht, 1981).
- <sup>26</sup>About 40 Mbytes of memory is necessary. Practically,  $N_p=50^3$  is sufficient and the computational time in this condition is 4.3 s and 1 h for  $l=2$  and  $l=10$ , respectively [a single CPU (PentiumIV EE) at 3.20 GHz].

This computational time depends strongly on  $N_p$  and  $l$  (when  $N_p=100^3$  is used, 32 s and 8 h for  $l=2$  and  $l=10$ , respectively). Our calculation of SDF by MD, which was performed with  $1 \times 10^6$  steps suggested by Ref. 4, takes about 10 h.

<sup>27</sup>S. Ten-no, F. Hirata, and S. Kato, J. Chem. Phys. **100**, 7443 (1994).

<sup>28</sup>H. Sato, F. Hirata, and S. Kato, J. Chem. Phys. **105**, 1546 (1996).

<sup>29</sup>P. Flükiger, H. P. Lüthi, S. Portmann, and J. Weber, MOLEKEL 4.0, Swiss Center for Scientific Computing, Manno, Switzerland, 2000.