Realization of Three-Dimensional Solvation-Structures from the Site–Site Radial Distribution Functions in Liquids

Hirofumi Sato* **and Fumio Hirata**

Department of Theoretical Studies, Institute for Molecular Science, and School of Mathematical and Physical Science,
The Graduate University for Advanced Studies, Okazaki 444-8585

(Received February 13, 2001)

We propose a new procedure to realize a three-dimensional (3D) solvation structure around a solute molecule from a set of radial distribution functions (RDF), or distance information. The method consists of the minimization of a penalty function defined as the mean-square difference of the solute–solvent interatomic distances obtained from trial 3D configurations and from target RDFs. The hydration structures around several different solute molecules are visualized to demonstrate the method.

The solvation structures realized with the present method correspond to the most plausible solvation structure (MPSS), which looks like a "snap shot" of the molecular dynamics trajectory. However, the MPSS is essentially different from the "snap shot," since it represents an average configuration, and is therefore an observable quantity. The present procedure was originally developed for analyses of the RDF results obtained from the reference interaction site model (RISM), but can be applied straightforwardly to RDFs from other sources, such as molecular simulations and scattering experiments.

The solvation structure around solute molecules is commonly described and understood in terms of the radial distribution function (RDF), representing a probability distribution of inter atomic distance. The function can be obtained by a computational approach (such as a molecular simulation and integral equation theory for a liquid)[1,2,3,4] as well as by scattering experiments. Although the RDF provides very useful information regarding the solvation structure, it is rather mathematical, and provides little chemical intuition.

Compared to a set of RDF, the distribution functions in three-dimensional (3D) space is more intuitive for understanding the chemical aspect of solvation. Unfortunately, it is generally formidable to characterize this type of function, since a huge number of pixels are required to figure out the profile of the functions. This is why even the orientation correlation function is rarely estimated in computer simulations. Several types of approaches have been proposed to obtain the 3D solvation structure from both experimental and theoretical sides. By imposing a maximum-entropy-like constraint, Soper generated orientational distribution functions via the standard spherical harmonic expansion from experimentally observed partial structure factors.[5] New types of integral equation theories for liquids, which provide a 3D distribution function, have been proposed by several researchers.[6,7,8,9] The 3D-RISM theory by Kovalenko et al.[8] has been combined with ab initio molecular orbital theory,[10] which provides 3D information about the solvation structure as well as the electronic distribution in solute molecules. However, computations in these theories are usually much more demanding than that of the traditional integral equation method based on RDF. Moreover, even the 3D distribution functions may still be far from chemical intuition due to their complicated shapes.

Considering these circumstances, a convenient procedure that extracts an intuitive 3D-picture of solvation from a set of RDFs, which is an alternative to the 3D distribution function, is desired. In the present article, we propose a simple model approximating 3D distribution functions for the solvation structure. Given a set of RDF, the method extracts a "cluster-like" solvation structure. The concept of a "cluster" is commonly employed among people who are trying to extend the approaches developed in the gas phase to the liquid phase in order to understand the solvation processes in solution. The present method is a reverse approach which abstracts a cluster picture from the solvation structure in solution. The solvation structure is viewed as a "cluster" fluctuating around its most plausible, or average, structure. The idea applied here is essentially the same as that is used in determing protein structures from a set of distance constraints obtained by the 2D-NMR method.[11,12] The present study is the first application to solute-solvent configurational problems, in which information about the distance constraints for all pairs between atoms is available.

The paper is organized as follows. The procedure used to construct MPSS is described in the following section. In order to verify the robustness of the present method, a molecular dynamics simulation was carried out for an aqueous solution of hydrogen fluoride, and the trajectory was analyzed in terms of both MPSS and the three-dimensional distribution functions projected into a contour map. In Results and Discussions, the method applied to the RDF results is discussed concerning solute-solvent systems, obtained previously from the RISM-SCF procedure. The paper is concluded in Section 4.

Method

Let's assume that a set of RDFs for all site (atom)-pairs between the solute and the solvent is provided from either experiments or theories. If one is interested in a system in which solute and solvent molecules have N_u and N_v sites, respectively, there are $N_u \times N_v$ RDFs to be considered. The geometry of the solute and solvent molecule is assumed to be known and to be rigid.

Our goal is to obtain a set of the distribution function $\{\rho_\gamma(\mathbf{r})\}$ for solvent site γ in 3D space from a set of the site–site RDFs. Here, we introduce a concept of the "most plausible solvation structure" (MPSS). Among various solvent configurations in a thermally averaged ensemble, the MPSS is the most frequently appearing one, which corresponds to the (local) maximum in the distribution function, $\rho_\gamma(\mathbf{r})$. Since we already have a set of RDFs (which can be obtained from RISM, molecular simulations and so on), we can construct a MPSS which is consistent to a set of the positions. (Step 1) We assume that the thermal fluctuations of the solvent configuration around the MPSS is described by Gaussian distributions in the sense of the central limit theorem.[13] Then, the distribution function, $\rho_\gamma(\mathbf{r})$, can be approximated as the superposition of harmonic fluctuations around the MPSS(Step 2),

$$\rho_\gamma(\mathbf{r}) \propto \exp\left\{-\beta\sum_i w_i(\mathbf{r}-\mathbf{r}_0^i)^\dagger\mathbf{K}^i(\mathbf{r}-\mathbf{r}_0^i)\right\}, \tag{1}$$

where $\{\mathbf{r}_0{}^i\}$ is a set of positions (MPSS), i runs over the solvent molecules consisting of MPSS, w_i is a weight function, $\beta = 1/k_BT$, and $\mathbf{K}^i$ is a real symmetric matrix representing the thermal fluctuation. It is noted that a projection of $\mathbf{K}^i$ on the vector between the α and γ sites can be computed with the second-order derivative of RDF around its peaks. Thus, two steps are contained to construct the 3D distribution function ($\rho_\gamma(\mathbf{r})$) in our approach. The first step is to determine the MPSSs and the second step is to estimate the thermal fluctuation around them.

The MPSS projected onto the RDF should correspond to its peaks, which means that the MPSS must be reconstructed essentially only from the information of the peak positions in RDFs. Thus, the MPSS is obtained by determining these six parameters (the position and Euler angle: $\mathbf{V}_O$, θ, ϕ and χ, see the Appendix) so as to reproduce these peak positions. With the information of the molecular geometry (Cartesian coordinates) of solute ($\{\mathbf{U}_j, j = 1,...,N_u\}$), the internal coordinate of the solvent and the RDFs between all pairs of the solute and solvent sites, we determine the solvent position ($\mathbf{V}_i$) by minimizing the following function that estimates the deviation error:

$$\chi_{\mathrm{SQ}} = \frac{1}{N_uN_v}\sum_{i,j}^{\text{all pair}}\{|\mathbf{V}_i - \mathbf{U}_j| - l_{ij}\}^2, \tag{2}$$

where i and j are sites in the solvent and solute molecules, respectively, and l_{ij} represents the radial distance between the two sites estimated from RDF. The minimized value of χ_{SQ} can be attached some physical meaning. If the value is small, the 3D structure so determined should reproduce the original set of RDFs reasonably well. In that sense, χ_{SQ} is a measure representing how "plausible" such a solvation structure or solvent configuration around the solute is. From such a consideration, we refer to χ_{SQ} as a "plausibility score".

In a practical treatment, the minimization is performed for individual solvent molecules varying the six parameters. We examined all of the solute-solvent site pairs (all possible sets of l_{ij}) up to the third peak, and solvent molecules with $\chi_{\mathrm{SQ}} \geq 0.5$ were ignored. In the present article, we report only on linear molecules in aqueous solution, though an extension to other solute-solvent systems is straightforward. It seems that the initial conditions for the minimization might be critical for the final results. The details of the choice in the present study are explained in Appendix.

The second step is to estimate the thermal fluctuations around the MPSS. The width of the peak in RDF is a manifestation of the fluctuations in the solvation structure. The well-known relationship between the potential of mean force ($w(r)$) and RDF,

$$w(r) = -k_BT \log g(r), \tag{3}$$

suggests that the fluctuation around the MPSS may be approximated by a Gaussian-type distribution if $w(r)$ is described by a harmonic potential. This is the simplest representation of fluctuations, which provides information about the solute-solvent binding strength. We examined a "force constant" (second order derivative), k, associated with the fluctuation of solvent sites (oxygen or hydrogen in the present article) by fitting RDF to a Gaussian-type function and by averaging over all of the solute sites. These quantities represent how loosely/tightly the solvent molecule is bound to the solute.

1 Verification by Molecular Simulation. In the present section, we compare MPSS with the distribution functions obtained from the molecular dynamics (MD) simulation to verify the robustness of the method in realizing the solvation structure. As an example, we consider the solvation structure around HF, which we studied previously by the RISM-SCF method.[14] Our MD model for HF and the solvent water is rigid with the same parameters used in the previous RISM-SCF study. A system consisting of 214 water molecules and one HF molecule located in a cubic simulation cell with periodic boundary conditions was examined for 1 ns, after a 1 ns equilibration run, at the same temperature and density as in the RISM-SCF study. ($\rho = 0.003334$ molecule/Å, $T = 298.15$ K) We computed the distribution functions on the plane containing the solute HF to obtain an accurate distribution, since it has been known that the convergence of the distribution function in high-dimensional space is not good.

In Fig. 1, the RDFs from MD are shown and compared with those from RISM-SCF. The peak positions and heights calculated with the two methods are in good qualitative agreement for the pair of solute (H and F) and water oxygen. The peak heights and positions of RDFs in solute and water hydrogen are also in reasonable agreement, but exhibit some difference, especially in the second peak position. The MPSS (three water molecules in the nearest neighbor) computed from MD RDFs[15] are shown in Fig. 2. As we pointed out in the previous study, there are two types of hydrogen bondings: one is between hy-

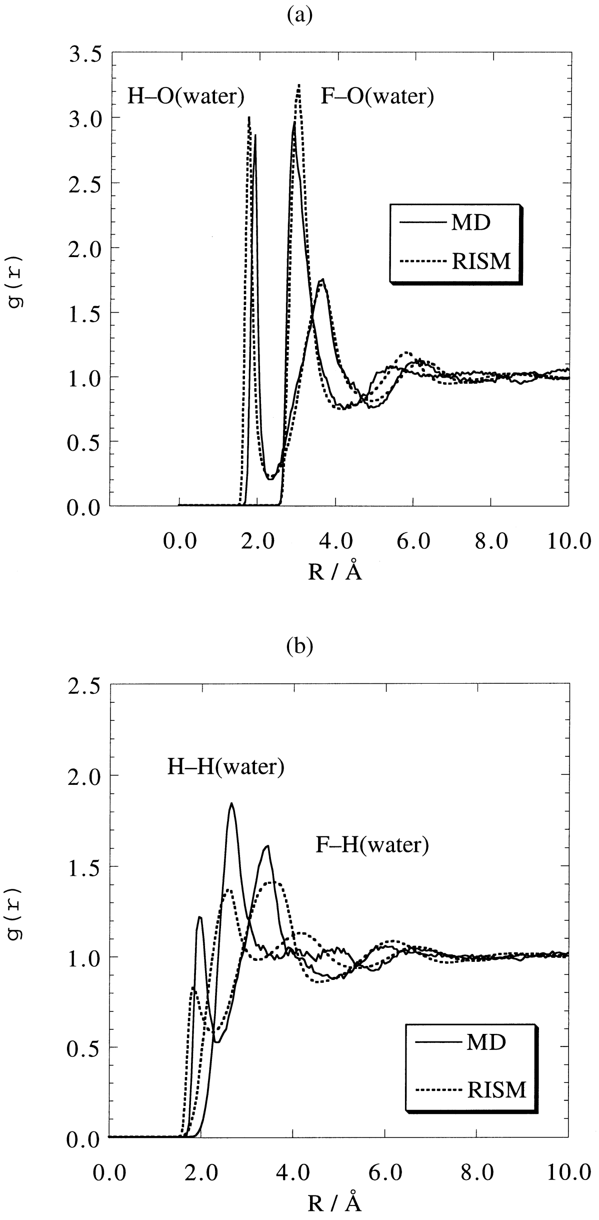

Fig. 1. RDFs between (a) solute sites (H and F) and water oxygen and (b) solute sites and water hydrogen, which are computed with MD simulation (solid lines) and RISM (dashed lines).

drogen (HF) and water oxygen; the other between fluoride and water hydrogen. In Fig. 3, the distribution functions computed from the MD trajectory are shown in contour maps. The points marked with circles (○) in the figure correspond to the atomic positions in MPSS obtained from the MD simulation. The important point to notice is that all of the points are located reasonably close to the local maximum in the distribution functions, clearly verifying that the present procedure can extract the most plausible configurations (or local maximum in the 3D distribution functions). It should be mentioned that the two points (MPSS) in the hydrogen atom distribution ($X \simeq -2$) are not very close to the local maximum, but are located on a broad hill. They can be assigned to the counter hydrogen atom

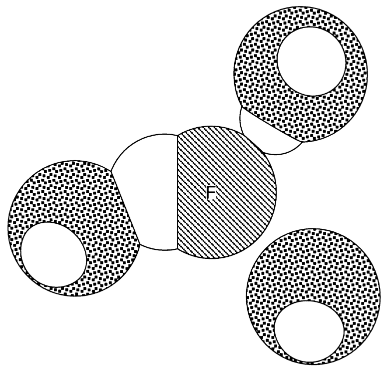

Fig. 2. MPSS (three solvent water molecules in the nearest neighbor) computed with RDF from MD simulation.

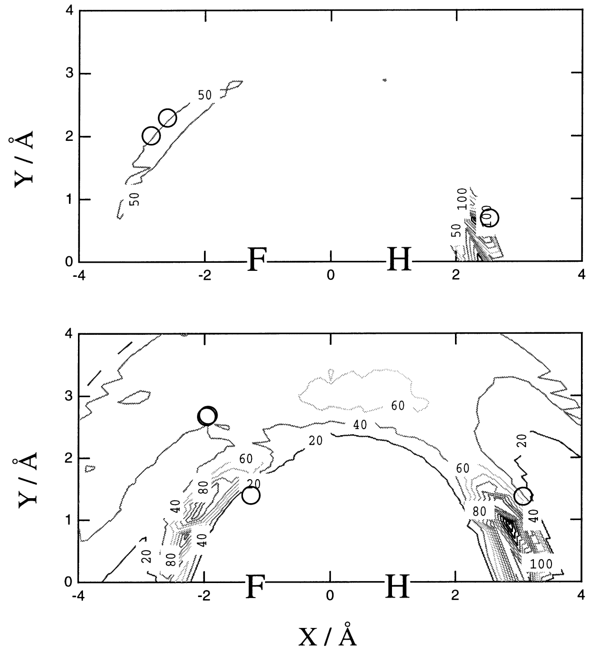

Fig. 3. Two dimensional distribution functions of water oxygen (the upper figure) and hydrogen (the lower figure). Unit for the distribution is arbitrary and normalized so as to 100 at the maximum point.

with the hydrogen-bonding atom ($X = -1.5$, $Y = 1.5$), and do not appear in the simulated distribution functions due to thermal averaging. It is obvious that MPSS can not represent a broad distribution, and this is an example of a limitation of the present method. However, there is no doubt that the MPSS is sufficient to realize a *qualitative* profile of the solvation structure.

It is worthwhile to emphasize at this moment that the MPSS analysis is more suited to RDFs obtained from the integral equation method than from molecular simulations and scattering experiments, because the latter two suffer from noise due to the statistical error. The noise sometimes obscures the positions of the peaks in RDF, especially those further than the second peak, and leads to incorrect results in MPSS. On the other hand, the RDF results from the RISM theory are free from such noise, and provide basically correct results for MPSS within the approximation associated with the theory. In what follows we present MPSS analyses for the RDF results from the RISM theory.

Results and Discussion

1 Carbon Monoxide. We choose carbon monoxide in aqueous solution as the first application of the MPSS analysis for the RISM RDFs, since the 3D distribution function is known from our previous study.[10] The RDFs were computed with the RISM-CASSCF method,[16] and consist of four unique pairs: C(solute)–O(solvent), C(solute)–H(solvent), O(solute)–O(solvent) and O(solute)–H(solvent). The resultant MPSS, the solute–solvent site distances in it, and the peak positions in the original RDFs are shown in Fig. 4. All three water molecules (W_1, W_2 and W_3) are located on the middle of CO carbon and oxygen sites, and their hydrogens are oriented outwards from the CO solute. These features are very similar to the 3D RISM-CASSCF results. As expected from a previous study, even a glimpse of hydrogen bonding between the CO carbon and hydrogen is not caught in the set of RDFs.[10]

We would like to draw attention to the following two points: The first concerns solvent molecules that are expected to be found near to the carbon or oxygen atoms along the direction of the CO molecular axis (shaded area shown in Fig. 5). The three water molecules in MPSS appear in the region perpendicular to the CO axis as drawn in the figure. This only indicates that there is strong correlation, positional as well as orientational, between the CO and the solvent molecules in that area. Of course, there are water molecules densely packed around the shaded region as well, although those molecules are not drawn explicitly. Those water molecules are randomly distributed positionaly as well as orientationaly with an average density of $\rho_\gamma(\mathbf{r}) = \rho$(bulk density), so that it does not give rise to MPSS.

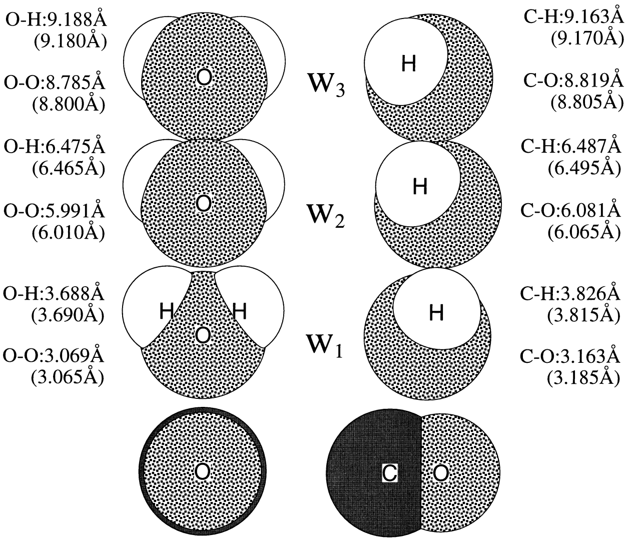

Fig. 4. The MPSS for carbon monoxide in aqueous solution viewed from oxygen atom direction and from the side. The values shown in the figure are the distance in the real 3D space (MPSS) compared to the peak positions appeared in RDF (given in the parenthesis).

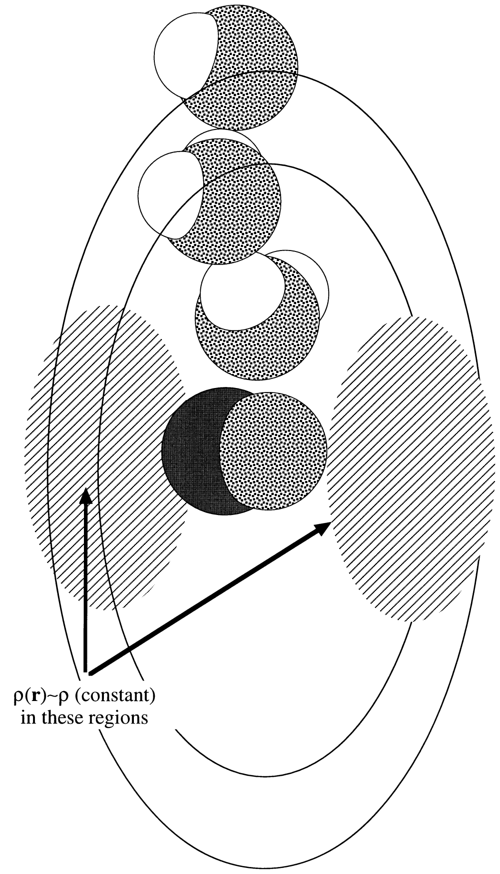

Fig. 5. Representation of cylindrical symmetry of linear molecules.

The other point to be made is that CO is a collinear molecule and there is an arbitrariness in determining the position of water molecules around the molecular axis. The distribution function in three-dimensional space, thereby, possesses cylindrical symmetry, although all of the water molecules in Fig. 4 are drawn in the same plane involving the CO axis. In the present procedure, the positions of water molecules are individually determined, and the correlation between the solvent molecules are not explicitly taken into account (these correlations are supposed to be included implicitly in the solution of the RISM equation). Figure 5 illustrates another version of the MPSS, in which the second and third water molecules are rotated by considering the explicit correlation between these molecules. What we have done is to search the rotational angle around the CO axis so as to reproduce the first and second peak positions of H–O (1.8 Å and 3.7 Å) and O–O (3.0 Å) in the RDFs of liquid water. The result suggests the existence of a hydrogen-bond between the first and second water molecules. This seems to be the origin of a rather strong correlation along the perpendicular direction of the CO axis. Nevertheless, the packing effect is the dominating factor to determine the solvation structure.

Table 1 shows the "force constants" (k) of the solute–solvent binding, the effective length of fluctuations ($R_{\mathrm{eff}} = \sqrt{k_B T / k}$) and their "vibrational frequencies" (ν), which were computed using the atomic mass of oxygen and hydrogen. One can see that the inner water molecules are bounded more tightly than

Table 1. Fluctuations of Water Molecules around Carbon Monoxide

Solvent Index	k/kcal (molÅ^2)$^{-1}$		ν/cm^{-1}		R_{eff}/Å*	
	O	H	O	H	O	H
W_1	2.50	0.54	43	79	0.49	1.05
W_2	0.26	0.08	14	31	1.51	2.66
W_3	0.04	0.02	5	15	4.02	5.72

*Computed with $R_{\text{eff}} = \sqrt{k_B T / k}$.

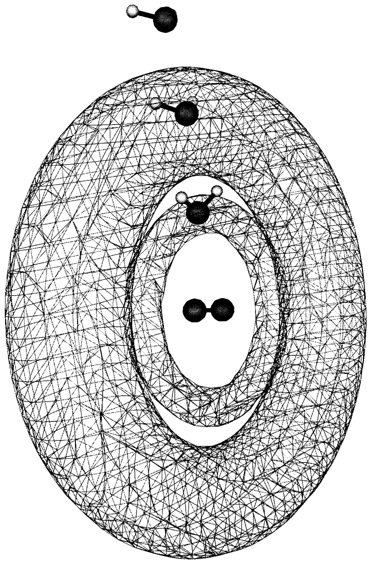

Fig. 6. The shape of solvent distribution function in three dimensional space. The wire frames represent the fluctuations estimated from the force constant listed in the Table 1.

the outer ones, and that the fluctuation of oxygen is smaller than that of hydrogen. It is noteworthy that these are quantities relative to those of the solute molecule, that is fixed in the 3D space. A more proper representation of the MPSSs for the linear solute molecule is exhibited in Fig. 6.

2 A Series of Hydrogen Halides. Recently, we investigated the acidity of hydrogen halides (HX) by means of the RISM-SCF method and clarified the origin of the unusually law acidity seen in hydrogen fluoride. The key to resolve the puzzling behavior is the number of hydrogen bondings between the solute in the undissociated form and water molecules. The number of hydrogen bonds is one for HCl, HBr and HI while two for HF.[14]

Figure 7 shows the MPSS of the four hydrogen halides in aqueous solutions. Two or three solvent molecules (W_1, W_2, and W_3) are seen along the H–X molecular axis in all cases, with the first one attached to the hydrogen atom of HX, suggesting the existence of a strong correlation along this direction.[17,18] As we have already discussed in the previous paper, there is hydrogen bonding between all the pairs of HX-hydrogens and water-oxygens. On the other hand, the halide side is not that simple in terms of the solute–water interaction. In all four cases, two water molecules (W_4 and W_5) are identified on the halide side, with the first one (W_4) attached to the halide-atom. However, only HF makes conspicuous hydrogen-bonding with a water molecule (W_0) on the fluoride-atom side (Fig. 7a).

Table 2 gives the "vibrational frequencies" of the MPSSs discussed above. The frequency concerning direct bindings or interactions between HX and water molecules, namely, W_0, W_1, and W_4, is relatively higher, and it gradually decreases in the outer water molecule (W_2). The table also shows a moderate tendency with respect to the dependency on the halide species. As the halide size becomes greater, the frequency generally becomes less. This is a manifestation of the electrostatic interaction between the solute and the solvent, which becomes weaker as the solute size becomes larger. On top of that, as already pointed out, only HF has an extra water molecule, W_0, hydrogen-bonded to the halide atom. The frequency is much higher (280 cm^{-1}) than that of the other hydrogen-bond between O(water) and H(HX), indicating that the extra-hydrogen bond is even stronger than the other.

3 Carbon Dioxide. The third example to be discussed here is a triatomic molecule in aqueous solution. It is observed empirically that the solubility of carbon dioxide in water is larger than that of CO, which apparently contradicts our naive intuition based on the free energy of cavity formation and classical interactions between those solutes with water molecules. In our previous study,[16] we found that hydrogen bonding between the CO_2 oxygen and water hydrogen is a key to understanding this puzzling behavior. Figure 8 shows the MPSS of CO_2 in aqueous solutions calculated from the RDFs obtained in the previous study based on the RISM-CASSCF method. Again, we would like to emphasize that there is a cylindrical symmetry in this system, and that the position of individual water molecules is arbitrary around the O–C–O axis. The hy-

Table 2. "Vibrational Frequencies" of Water Molecules around HX

	HF			HCl			HBr			HI		
	ν_O/cm^{-1}	ν_H/cm^{-1}	ν_H/cm^{-1}	ν_O/cm^{-1}	ν_H/cm^{-1}	ν_H/cm^{-1}	ν_O/cm^{-1}	ν_H/cm^{-1}	ν_H/cm^{-1}	ν_O/cm^{-1}	ν_H/cm^{-1}	ν_H/cm^{-1}
W_0	58	280	49									
W_1	143	120	120	105	104	104	86	85	85	62	—	—
W_2	29	47	47	17	39	39	19	37	37	18	35	35
W_3	—	—	—	6	22	22	6	19	19	6	16	16
W_4	58	49	49	43	61	61	42	57	57	39	51	51
W_5	18	39	39	14	32	32	14	30	30	14	28	28
W_6	18	49	39									

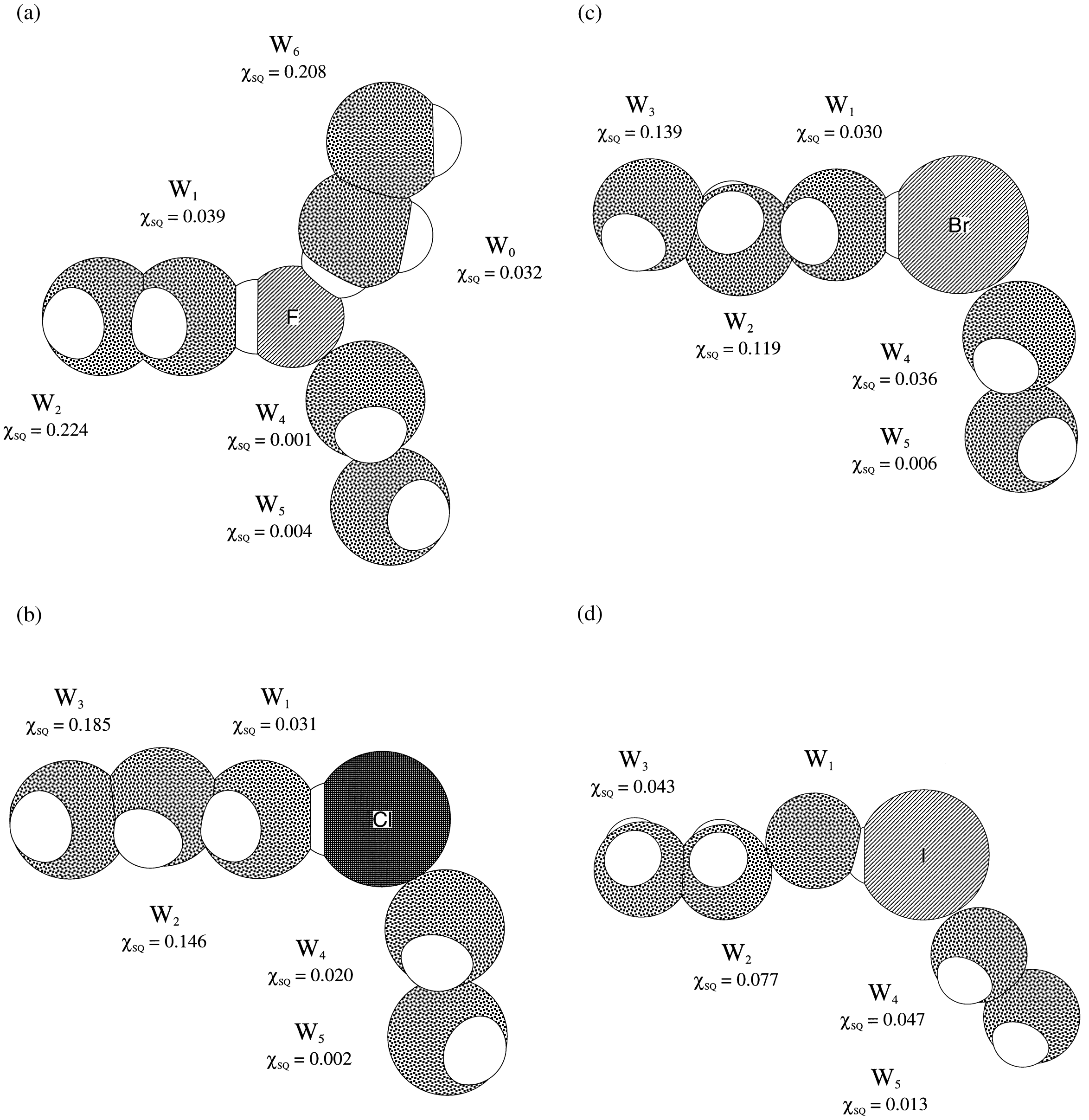

Fig. 7. The MPSS for a series of hydrogen halides (HX) in aqueous solution. (a) HF (b) HCl (c) HBr (d) HI. Note that there is cylindrical symmetry in all the case of HX and all the solvent configuration generated by arbitrary rotations around HX axis is equivalent.

drogen-bonding mentioned above is formed with one of the oxygen atoms of CO_2, which is labeled as W_3 in the figure. At the same time, a water molecule, W_4, which is not hydrogen-bonded, is identified at the other oxygen of CO_2. These water configurations are equally plausible because they have a similar plausibility score, χ_{SQ}. Such a plausible configuration can be explained by a packing effect or a volume exclusion effect.

Water molecules seem to access to the central carbon atom directly along the direction perpendicular to the O–C–O axis (W_1 and W_2). The former water molecule shows a specific hydrogen coordination, but this type of solvation seems not to be so plausible because of the large value of χ_{SQ}. It is therefore natural to conclude that the dominant configuration of water molecules at the carbon atom of CO_2 must be W_2. This example emphasizes the importance of the plausibility score, χ_{SQ}, for realizing the most plausible structure.

Table 3 lists the "vibrational frequencies" of MPSSs. A trend similar to that found in the case of hydrogen halides can been seen for water molecules in the different coordination shells. The frequency is generally higher for water molecules in direct contact with the solute atoms, and it gradually reduces in the outer coordination shells. It is readily noticed that two water molecules, W_1 and W_3, which are in direct contact with the solute, have a large frequency (247 and 185 cm^{-1}). The high frequency associated with W_3 is easily understood since it forms a hydrogen bond with the oxygen atom of the solute. However, the large frequency of W_1 is rather odd, because it should not be making any hydrogen bond with the carbon atom

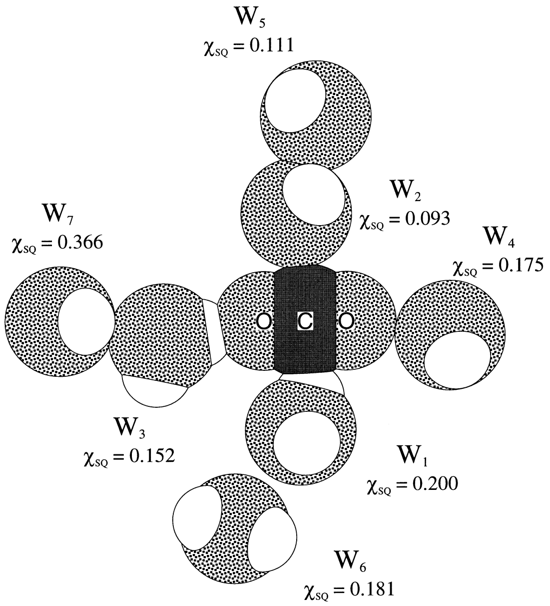

Fig. 8. The MPSS for carbon dioxide in aqueous solution. Note that there is cylindrical symmetry. See the footnote in Fig. 7.

Table 3. “Vibrational Frequencies” of Water Molecules around CO_2

Solvent Index	ν_O / cm^{-1}	ν_H / cm^{-1}	ν_H / cm^{-1}
W_1	47	247	87
W_2	47	87	87
W_3	35	185	65
W_4	35	65	65
W_5	13	32	32
W_6	10	32	32
W_7	7	32	32

of the solute, as can be figured out from its large χ_{SQ} value. We believe that it is an artifact caused by the fact that the two water molecules, W_1 and W_3, give rise to the similar distance between $O(CO_2)$ and $H(H_2O)$. As discussed in the previous study, the peak profile of RDFs between water and CO_2 carbon/oxygen are broad, indicating that the solvent molecules distribute continuously from the carbon site to the oxygen site of the solute CO_2. In other words, the method can not distinguish the two water molecules, W_1 and W_3.

Concluding Remarks

In the present article, we propose a new procedure to realize three-dimensional solvation structures around a solute molecule from the site–site radial distribution functions. The computational results of RISM are expressed in a set of one-dimensional RDFs that may be slightly different from an *intuitive* understanding of the solvation structure. The new procedure provides a direct way to connect the RISM results with the visible molecular solvation profile, which we refer to as the most plausible solvation structure (MPSS).

The MPSS looks similar to a “snap shot” taken from a trajectory of a molecular dynamics simulation. Both of them give us an geometrical picture of solvation that helps our intuitive understanding of chemical phenomena in solution. However, they are completely different in the following sense. A snap shot from the molecular simulation is one of the solvation structures sampled from thermally fluctuating configurations, which may or may not be a most plausible structure. The MPSS, on the other hand, is a three-dimensional realization of a structure averaged over such fluctuating configurations. In that sense, the MPSS presented in this paper is an observable, while a snap shot from the molecular simulation is not.

One may be aware of a method called “distance geometry,” or the metric matrix method, which was developed in the field of protein chemistry in order to find the 3D structure of a biomolecule from distance information provided by the 2D-NMR.[19] This method can also be applied to find MPSS from RDF information. We discuss the relationship between the present method and the metric matrix method in a forthcoming report.

Although only systems with collinear solute molecules in aqueous solution are discussed in the present article, an extension of the method to a more general class of solute/solvent systems is straightforward. It is noted that if the number of sites in a solute molecule is greater than four, the ambiguities in determining the solvent position must diminish. An application of this method to solute molecules with larger size is in progress.

Appendix

Description of Solvent Position. The solvent position is completely determined by six coordinates: three for the position and three for the orientation. In the case of water as a solvent, the positions of two hydrogens ($\mathbf{V}_H$) are expressed as follows:

$$\mathbf{V}_H = \mathbf{V}_O + D(\mathbf{e}_P, \mathbf{e}_Q, \mathbf{e}_R)\cdot\begin{pmatrix}\sin\theta\cos\phi\\ -\sin\theta\sin\phi\\ \cos\theta\end{pmatrix}$$
$$\pm\, r(\mathbf{e}_P, \mathbf{e}_Q, \mathbf{e}_R)\cdot\left\{\left[\begin{pmatrix}-\sin\phi\\ \cos\phi\\ 0\end{pmatrix}\cos\chi + \begin{pmatrix}-\cos\theta\cos\phi\\ -\cos\theta\sin\phi\\ \sin\theta\end{pmatrix}\sin\chi\right]\right\}, \tag{4}$$

where $\mathbf{V}_O = V_O^x, V_O^y, V_O^z)$ is the position of the oxygen atom, $\mathbf{e}_P$, $\mathbf{e}_Q$, and $\mathbf{e}_R$ are an arbitrary set of normalized orthogonal vectors, D represents the length of the line bisecting the H–O–H angle from the oxygen to the orthocenter on H–H, and r is half the distance between the two hydrogen atoms (Fig. 9A1).

Peak Position and Cartesian Coordinate. Before performing the χ_{SQ} minimizing procedure, we have to make good initial guess concerning the positions of solvent molecules, which are specified by the six parameters defined by Eq. 4. The initial structure in 3D space is given by the following procedure with the interatomic-distance information included in the site–site RDFs.

Among the N_u RDFs between a site in a solvent molecule and all of the solute sites, the first and second highest peaks are chosen. We define two position vectors ($\mathbf{P}_1$ and $\mathbf{P}_2$) of the solute sites contributing to these two peaks, which appear at distances of l_1 and l_2 in each solute-solvent pair of RDF. Then, the points that are located at distances of l_1 and l_2 from $\mathbf{P}_1$ and $\mathbf{P}_2$ would be a circle in

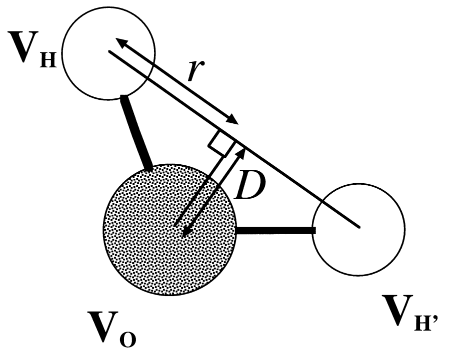

Fig. 9. Fig. A1. Modeling of solvent water.

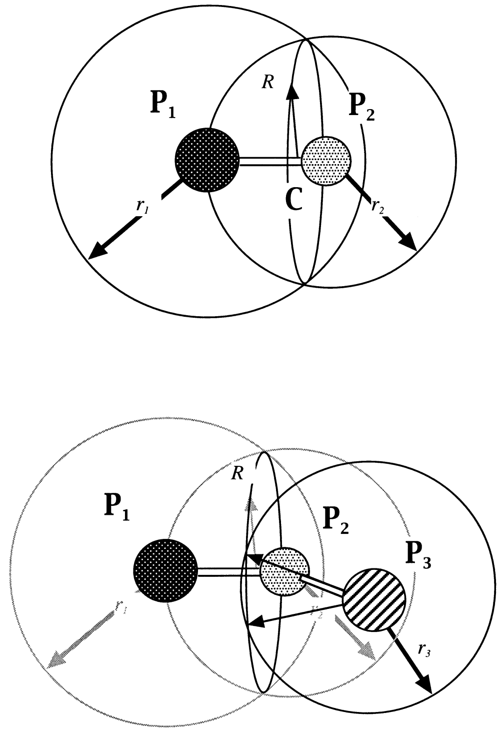

Fig. 10. Fig. A2. Relationship between Cartesian coordinates and peak positions.

3D space (see Fig. 10A2). The center (**C**) and radius (R) of the circle are

$$\mathbf{C} = \mathbf{M} + \frac{l_1^2 - l_2^2}{2|\mathbf{P}|^2}\mathbf{P},$$

$$R^2 = \frac{l_1^2 + l_2^2}{2} - \frac{|\mathbf{P}|^2}{4} - \frac{(l_1^2 - l_2^2)^2}{4|\mathbf{P}|^2},$$

where

$$\mathbf{P} = \mathbf{P}_2 - \mathbf{P}_1,$$

$$\mathbf{M} = \frac{\mathbf{P}_1 + \mathbf{P}_2}{2}.$$

If the solute molecule consists of more than three sites, there is another position vector ($\mathbf{P}_3$) which is related to the third-highest peak at the position of l_3. Now, we introduce two normal vectors, $\mathbf{e}_Q$ and $\mathbf{e}_R$; $\mathbf{e}_Q$ is perpendicular to $\mathbf{e}_P = \mathbf{P}/|\mathbf{P}|$ and containing the point of $\mathbf{P}_3$, while $\mathbf{e}_R$ is perpendicular to both $\mathbf{e}_P$ and $\mathbf{e}_Q$. The solvent position (**V**) that is consistent with the RDF peak positions (l_1, l_2, l_3) is given by

$$\mathbf{V} = \mathbf{C} + R\,(\mathbf{e}_Q\cos A + \mathbf{e}_R\sin A),$$

$$\cos A = \frac{|\mathbf{Q}'|^2 + R^2 - l_3^2}{2R\mathbf{Q}'\cdot\mathbf{e}_Q},$$

where

$$\mathbf{Q}' = \mathbf{P}_3 - \mathbf{C}.$$

In the case of a collinear solute molecule, the choice of angle A is arbitrary due to the high symmetry. Note that the sign of $\sin A$ can be determined if the forth-highest peak is given, because a triatomic molecule always possesses planner symmetry.

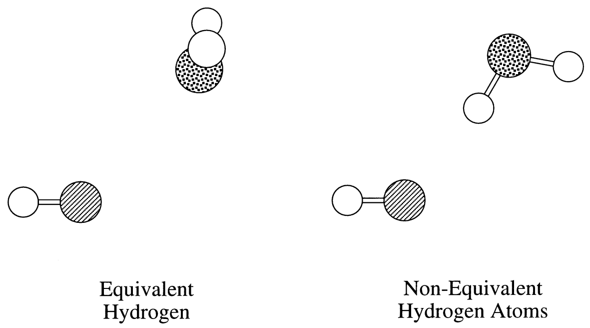

Fig. 11. Fig. A3. Two possible orientational configurations of water molecules.

Molecular Orientation The orientation of a water molecule is determined by the Euler angle or by the positions of the two hydrogens. Since the Euler angle contains three parameters, information of RDF between the two water hydrogen atoms and the two solute sites (peak positions in the four pairs of RDF) is sufficient to make an initial guess. We thus assume two types of molecular orientations with respect to the two solute sites ($\mathbf{P}_A$, $\mathbf{P}_B$), as the followed (see Fig. 11A3).

One is coordination with two hydrogen atoms being equivalent to the solute sites. With setting $= \pi/2$ and $\theta = \pi/2$ in Eq. 4, the description of the two positions of hydrogen atoms ($\mathbf{V}_H$) is reduced to

$$\mathbf{V}_H = \mathbf{V}_O + D\,(\mathbf{e}_P\cos\phi + \mathbf{e}_Q\sin\phi) \pm r\,\mathbf{e}_R.$$

By using information concerning the peak position in the RDFs between the hydrogen atom and site "A" (l_{H-A}) and "B" (l_{H-B}), the initial value is

$$\cos\phi = \frac{1}{2D|\mathbf{P}_B - \mathbf{P}_A|}\{(l_{H-A}^2 - l_{H-B}^2) + |\mathbf{V}_O - \mathbf{P}_B|^2 - |\mathbf{V}_O - \mathbf{P}_A|^2\}.$$

The other is the case in which the five atoms (two atoms in solute and three in water) are in the same plane. $\chi = 0$ and $\theta = \pi/2$ are set in Eq. 4, giving rise to

$$\mathbf{V}_H = \mathbf{V}_O + D\,(\mathbf{e}_P\cos\phi + \mathbf{e}_Q\sin\phi) \pm r\,(-\mathbf{e}_P\sin\phi + \mathbf{e}_Q\cos\phi),.$$

$$\cos\phi = \frac{1}{2D|\mathbf{P}_B - \mathbf{P}_A|} \left\{ \left(\frac{l_{H1-A}^2 + l_{H2-A}^2}{2} - \frac{l_{H1-B}^2 + l_{H2-B}^2}{2} \right) + |\mathbf{V}_O - \mathbf{P}_B|^2 - |\mathbf{V}_O - \mathbf{P}_A|^2 \right\}.$$

Although the initial vales are set with respect to only two atoms in solute molecules, a suitable orientation must be achieved during the minimization procedure, even for a triatomic (or more larger) solute molecule.

This research was supported by the Grant-in Aid for Scientific Research on Priority Areas "Molecular Physical Chemistry" (403-11166276), Grant-in Aid for Encouragement of Young Scientists (12740329), and from the Ministry of Education, Science, Sports and Culture.

References

1 D. Chandler and H. C. Andersen, *J. Chem. Phys.*, **57**, 1930 (1972).

2 F. Hirata and P. J. Rossky, *Chem. Phys. Lett.*, **83**, 329 (1981); F. Hirata, B. M. Pettitt, and P. J. Rossky, *J. Chem. Phys.*, **77**, 509 (1982); F. Hirata, P. J. Rossky, and B. M. Pettitt, *J. Chem. Phys.*, **78**, 4133 (1983).

3 S. Ten-no, F. Hirata, and S. Kato, *J. Chem. Phys.*, **100**, 7443 (1994).

4 H. Sato, F. Hirata, and S. Kato, *J. Chem. Phys.*, **105**, 1546 (1996).

5 A. K. Soper, *J. Chem. Phys.*, **101**, 6888 (1994); T. Yamaguchi and A. K. Soper, *J. Chem. Phys.*, **110**, 3529 (1999).

6 M. Ikeguchi and J. Doi, *J. Chem. Phys.*, **103**, 5011 (1995).

7 D. Beglov and B. Roux, *J. Phys. Chem. B*, **101**, 7821(1997).

8 A. Kovalenko and F. Hirata, *Chem. Phys. Lett.*, **290**, 237 (1998).

9 D. Beglov and B. Roux, *J. Chem. Phys.*, **103**, 360 (1995).

10 H. Sato, A. Kovalenko, and F. Hirata, *J. Chem. Phys.*, **112**, 9463 (2000).

11 G. M. Crippen and T. Havel, *Acta Crystallogr. Sect. A*, **34**, 282 (1978).

12 W. Braun and N. Go, *J. Mol. Biol.*, **186**, 611 (1985).

13 M. Toda, R. Kubo, and N. Saito, "Statistical Physics," Springer-Verlag, Berlin Heidelberg 1983 (1992).

14 H. Sato and F. Hirata, *J. Am. Chem. Soc.*, **121**, 3460 (1999).

15 Before the MPSS generation, the RDFs obtained from MD are smoothed to eliminate noises come from statistical errors.

16 H. Sato, N. Matubayasi, M. Nakahara, and F. Hirata, *Chem. Phys. Lett.*, **323**, 257 (2000).

17 For HF, there is additional peak at inner distance (1.8 Å) in F–H RDF that represents extra hydrogen bonding. Since the present computations were restricted upto the third peak in each RDF, we did not confirm the third solvent molecule for HF system, corresponding to the index of W_3.

18 The H–H RDFs of HI shows a shoulder around 2.7 Å,[14] which is arisen from the hydrogen bond between HI hydrogen and solvent oxygen. Since the present method searches the peak in RDF, we can not fix the position of hydrogen of this water molecule (Fig. 7d).

19 J. M. Blaney and J. S. Dixon, "Reviews in Computational Chemistry," VCH Publishers, Inc. New York (1994), Volume V, Chap. 6.