

Effect of Available Volumes on Radial Distribution Functions

TIMOTHY ASTLEY,¹ GORDON G. BIRCH,² MICHAEL G. B. DREW,¹
P. MARK RODGER,¹ GARETH R. H. WILDEN¹

¹Chemistry Department, University of Reading, Whiteknights, Reading, United Kingdom RG6 6AD

²Department of Food Science and Technology, University of Reading, Whiteknights, Reading, United Kingdom RG6 6AP

Received 11 December 1996; accepted 8 September 1997

ABSTRACT: The traditional method of analyzing solution structuring properties of solutes using atom–atom radial distribution functions (rdfs) can give rise to misleading interpretations when the volume occupied by the solute is ignored. It is shown by using the examples of O(4) in α - and β -D-allose that a more reliable interpretation of rdfs can be obtained by normalising the rdf using the available volume, rather than the traditional volume of a spherical shell. 1998 John Wiley & Sons, Inc. *J Comput Chem* 19: 363–367, 1998

Keywords: solution structuring; radial distribution functions; available volume

Atom–atom pair correlation functions, or radial distribution functions (rdfs), are a common tool used in molecular dynamics studies to describe the distribution of solvent molecules around a solute. They have the advantage of being directly related to observable experimental quantities¹—neutron and X-ray scattering intensities—although the full set of atom–atom rdfs is often difficult to extract from experiments on molecular systems. A great deal of information has been obtained from the position, intensity, and width of the various peaks exhibited by rdfs. For example,

in studies with water as the solvent, the intensity of the first peak in an rdf around a specific solute atom is often interpreted in terms of the hydrogen bonding character of that atom: A sharp peak in the O–O rdf at a radius of ca. 2.7 Å indicates a strong hydrogen bond, whereas a less well-defined peak structure with the first maximum beyond 3 Å is taken to indicate a hydrophobic interaction. Any residual structure beyond the first maximum is then used to probe the degree of long-range ordering of the solvent induced by the solute, and particularly by the atom or functional group on which the rdf is centered. The water structuring behaviours of α -D-glucose,² acetonitrile,³ dimethylsulfoxide (DMSO),^{4,5} methane,⁶ methanol,⁷ acetone,⁷ and ammonia^{6,7} are among many that have been interpreted in this manner.

Correspondence to: T. Astley

Contract/grant sponsors: BBSRC (U.K.); EC/AIR;

Contract/grant number: PL 94-2017

The purpose of this paper is to point out that such interpretations can be seriously misleading. The rdf is defined in such a way that deviations from a value of unity should indicate structure, with values greater than one indicating a local accumulation of atoms and values below one indicating a partial void. In other words, the rdf is the ratio of the number of atoms found in a given region to the average number one would expect to find if that region exhibited the bulk density. However, the rdf is defined as a spherical average, so this normalization* is achieved by comparison with the net number of particles found in a thin spherical shell with the bulk density. For solute-solvent rdfs this choice of normalization is often inappropriate, insofar as the excluded volume of the solute will cause a nonspherical distribution of solvent around individual solute atoms before any solvent structuring is even considered. In practical terms this means that solvent is excluded from part of the thin spherical shells used in normalizing the rdfs, so the conventional method will overestimate the normalization factor. A more reasonable procedure would be to use that part of each spherical shell that is available to the solvent in normalization, and we shall refer to this as the *available volume* method. Another source of potential error is in using rdfs as indicators of structural change in mixtures of different compositions. Vaisman and Berkowitz⁵ have argued that, in a mixture, an enhancement in the peak intensity relative to that of the pure liquid may result from a reduction in the solvent average density when the solute concentration increases. Thus, they suggest, an enhancement in the $g(r)$ peak heights does not necessarily mean an increase in structural order. One consequence of overlooking this effect is that comparison of the solvent structure around different solutes is likely to be misleading unless average volumes are considered. With a single solute of size and shape comparable to those of water³⁻⁵ or a group of solutes of similar size and shape,^{6,7} the effect may be negligible, but with solutes such as glucose, and especially disaccharides, the results can be very misleading, as we show in this paper.

Herein we illustrate the effect of using the available volume instead of the notional spherical vol-

*The word *normalization* has also been used in this context previously^{8,9} such that the normalization of an rdf ensures that $g(r \rightarrow \infty) = 1$ by comparing the observed density within the spherical shell to that predicted for a homogeneous system. By *normalization factor* we therefore mean the number of atoms expected to be found in the differential volume for a homogeneous distribution.

ume to normalize the rdf. We present two cases, one in which an apparently hydrophobic site is shown to be hydrophilic and one in which the use of available volumes identifies secondary structure that was not readily apparent using the traditional method.

In terms of the finite shells used in simulations, the rdf may be defined as

$$g_{\alpha\beta}(r) = \frac{N_{\alpha\beta}(r, \Delta r)}{\rho_{\beta} V_s(r, \Delta r)},$$

where $N_{\alpha\beta}(r, \Delta r)$ is the average number of β sites located in the shell of radius r and thickness Δr centered on site α ; $\rho_{\beta} = N_{\beta}/V$ is the average number density of β sites in the system, where N_{β} is the number of β sites in the system: $V_s(r, \Delta r) = 4\pi r^2 \Delta r$, the volume of a thin shell at radius r and thickness Δr . The use of available volume now proceeds simply by replacing the volume of the shell, V_s , with the volume of that shell available to the solvent, V_a .

The factor $\rho_{\beta} V_a(r, \Delta r)$ can be calculated using a Monte Carlo integration. A solute molecule was placed in an empty box of the same size as that used in the molecular dynamics simulation being investigated. A trial insertion of a water molecule at a random¹⁰ location within the box was then made. The trial insertion was rejected if the water molecule overlapped the solute molecule, where the overlap was defined on the basis of a spherical excluded volume about each solute atom. For our purposes, the radius of the excluded volume for a given solute atom α and solvent β was defined to be the largest distance for which the conventional $g_{\alpha\beta}(r)$ was zero in the full molecular dynamics simulations; i.e., the excluded volume actually observed in the molecular dynamics simulations. Typical values are given in Table I. The insertion process was then repeated a number of times *without* checking for solvent-solvent overlap; i.e., the location of each new water molecule was not dependent on the location of previous insertions. The

TABLE I. Radius of Exclusion for the Various Atom Types Found in Monosaccharides.

Atom type	Radius of exclusion (Å)
Hydroxyl-type oxygen	2.3
Carbon	2.5
Ether-type oxygen	2.5
Hydrogen	1.8

available volume was then calculated from

$$V_a(r, \Delta r) = \frac{n_s}{n_a} \times V,$$

where n_s = number of successful insertions into a shell of radius r and thickness Δr centered on atom α , n_a = total number of attempted insertions (10 times the number of waters in the simulation times the number of configurations studied), and V is the volume of the simulation box. To obtain a low level of statistical noise, the number of insertion attempts made was 10 times the number of solvent molecules in the simulation box. The number of insertion attempts was chosen as 10 after considering the balance between cpu demands, there being a linear increase with respect to the number of insertions, and accuracy. The accuracy was calculated by fitting a sixth-order polynomial through the available volume calculated for the first 75 histogram bins and calculating the average of the square of the deviation between the two. Table II shows that beyond 10 there is no further significant decrease in "noise." For preliminary calculations we use a factor of three; which gives an adequately low level of noise, while being low on cpu usage, and other choices can be made as the situation warrants. Finally, the new rdf was calculated using V_a .

The molecular dynamics package DL_POLY¹¹ was used to perform the simulations. In all cases the system consisted of a single solute atom surrounded by 241 water molecules in a cubic box of length 20 Å and subject to periodic boundary conditions to eliminate edge effects. A timestep of 0.5 fs was used. The potential used was a typical CHARMM-type molecular mechanics force field¹² for the sugar and the SPC¹³ water model for the solvent. The method is documented in full elsewhere.¹⁴

In the present study we have concentrated on monosaccharides in water, the rdf around each oxygen atom has been used¹⁰ to determine the

TABLE II. Standard Deviation of Available Volume Curve from Sixth-Order Polynomial of Best Fit

Number of insertions	Standard deviation
3	0.00119
5	0.00066
10	0.00045
20	0.00044

degree and strength of the hydrogen bond formation between the solute and water. Owing to the shape of a monosaccharide molecule, there can be a substantial amount of excluded volume (and hence little available volume for the solvent) around each oxygen atom, and this bulk is not spherically distributed, giving rise to spurious structure in the rdfs. More importantly for comparing oxygens within the same molecule, the available volume around each atom is not identical. Figure 1 shows the distribution of the ratio between the available and spherical volumes of the five hydrogen bonding oxygens of α -D-allose (shown in Fig. 2a). O(6) has more volume available to it at all distances below 4 Å, in keeping with the fact that it is part of the pendant CH₂OH moiety and thus has more "space" around it. O(4) is more confined than the other oxygens at shorter distances, because of the presence of this neighbouring CH₂OH group.

The result of using the available volume is presented in Figure 3, which shows the rdf around O(4) in β -D-allose (shown in Fig. 2b). The conventional rdf shows the standard hydrophilic interaction between the solute atom and the solvent, with minimal additional structure beyond 4.5 Å. However, once the excluded volume is taken into account, the first peak at 2.6 Å increases in intensity quite dramatically and a second peak showing secondary structure around 4.75 Å becomes much more obvious. Thus, by accounting for available volumes, solvent structure is elucidated that would otherwise not be apparent.

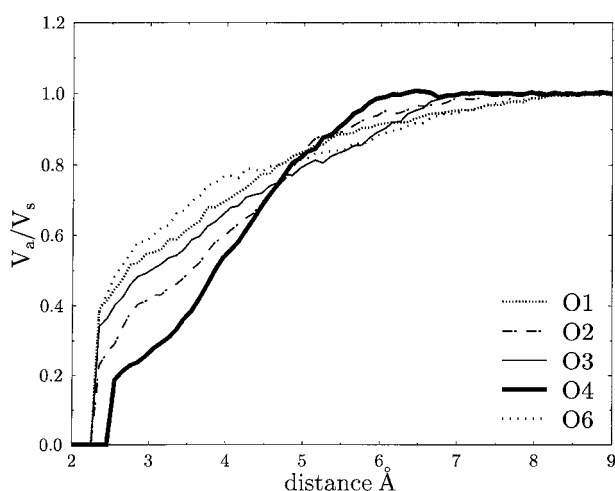


FIGURE 1. Plot of the ratio of available volume (V_a) to spherical volume (V_s) against distance (Å) for the five hydrogen bonding oxygens of α -D-allose.

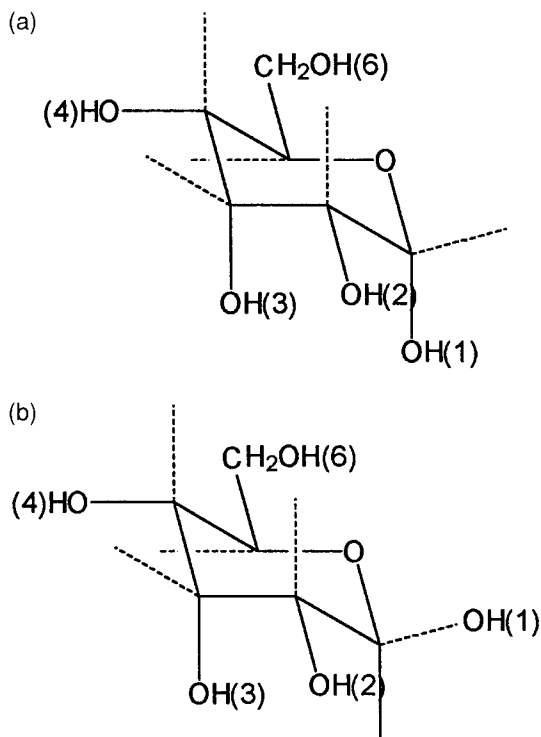


FIGURE 2. (a) α -D-allose. (b) β -D-allose.

Another example can be found for the rdf of O(4) in α -D-allose (Fig. 4). Here, the traditional rdf is typical of a site in a predominantly hydrophobic environment (though it should be noted that some O—O contacts as small as 2.5 Å are observed, which are indicative of the presence of hydrogen bonds). Using the available volume method, however, it becomes clear that there is strong H-bonding, but with access between the solvent and

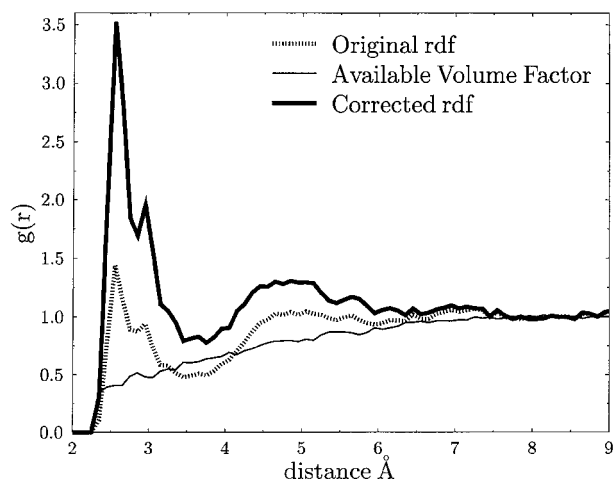


FIGURE 3. Traditional and available volume radial distribution functions for O(4) in β -D-allose.

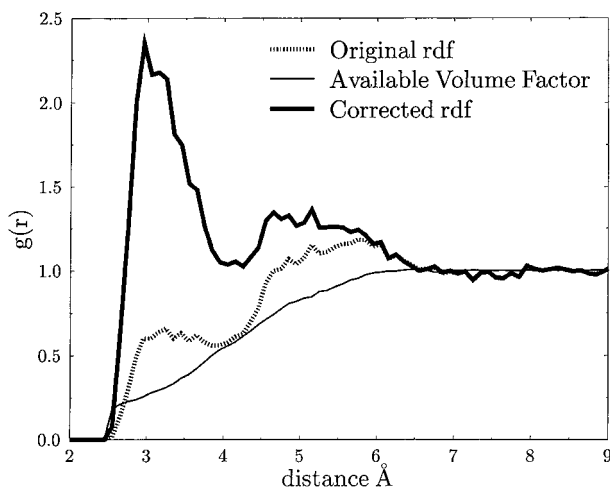


FIGURE 4. Traditional and available volume radial distribution functions for O(4) in α -D-allose.

solute limited by the presence of the CH_2OH group. Secondary structure, denoted by the presence of a peak in the rdf at 5 Å, is more akin to that usually observed for hydrophilic oxygens in monosaccharides than that found using the traditional method. It would be wrong to say that this oxygen is just as "solvated" as an oxygen with a standard rdf, because it clearly does have fewer oxygens around it. The available volume method demonstrates that the difference in the traditional rdf from what is expected for a carbohydrate OH is due primarily to steric hindrance and not to a change in the hydrophilicity of the oxygen.

By using available volumes within spherical shells to normalize rdfs, it is possible to gain a more reliable interpretation of solvent structuring effects. Two examples have been presented of cases in which proper accounting of the available solvent volume identifies structure that would otherwise have been overlooked. In an extreme case, the neglect of the available volume effect could lead to the incorrect categorization of hydrogen bonding sites as being hydrophobic. The traditional rdf still has a role in measuring the number of water molecules around a site and for comparison with experimental results, but it should be used in conjunction with the available volume calculations to ascertain the nature of the bonding to avoid misleading interpretations.

References

1. C. Andreani, J. C. Dopre, and F. P. Ricci, *Rep. Progr. Phys.*, **54**, 731 (1981).

2. J. W. Brady, *J. Am. Chem. Soc.*, **111**, 5155 (1989).
3. H. Kovacs and A. Laaksonen, *J. Am. Chem. Soc.*, **113**, 5596 (1991).
4. A. Luzar and D. Chandler, *J. Chem. Phys.*, **98**, 8160 (1993).
5. I. I. Vaisman and M. L. Berkowitz, *J. Am. Chem. Soc.*, **114**, 7889 (1992).
6. I. I. Vaisman, F. K. Brown, and A. Tropsha, *J. Phys. Chem.*, **98**, 5559 (1994).
7. M. Ferrario, M. Haughney, I. R. McDonald, and M. L. Klien, *J. Chem. Phys.*, **93**, 5156 (1990).
8. D. C. Rapaport, *The Art of Molecular Dynamics Simulation*, Cambridge University Press, Cambridge, U.K., 1995.
9. M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, Oxford, U.K., 1987.
10. S. K. Park and K. W. Miller, *Commun. ACM*, **31**, 1192 (1988).
11. W. Smith and T. R. Forester, DL_POLY, DRAL Daresbury Laboratory, Daresbury, Warrington, WA 4 4AD, U.K.
12. CHARMM, Version 3.2, Molecular Simulations Inc., 200 Fifth Avenue, Waltham, MA 02154.
13. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, In *Intermolecular Forces*, B. Pullman, ed., Reidel: Dordrecht, The Netherlands, p. 331 (1981).
14. T. Astley, G. G. Birch, M. G. B. Drew, P. M. Rodger, and G. R. H. Wilden, *Food Chem.*, **56**, 231 (1996).