PHYSICAL REVIEW B VOLUME 60, NUMBER 17 1 NOVEMBER 1999-I

Partial pair correlation functions of liquid water

László Pusztai
Research Institute for Solid State Physics and Optics, Hungarian Academy of Sciences, P.O.B. 49, H-1525 Budapest, Hungary

(Received 23 April 1999)

A procedure is proposed for deriving the partial pair correlation functions (PPCF's) of liquid water from the experimental total structure factors, applying the reverse Monte Carlo modeling technique [R. L. McGreevy and L. Pusztai, Mol. Simul. **1**, 359 (1988)]. The analyses of data on ambient liquid water lead to sets of PPCF's which differ drastically from the most widely accepted set of Soper, Bruni, and Ricci [J. Chem. Phys. **106**, 247 (1997)]. The most important difference is that the H-bonding distance is always bigger than suggested previously, at least by about 0.1 Å. [S0163-1829(99)00538-X]

To know the structure of a multicomponent liquid, like water, it is essential to know at least all of its partial pair correlation functions (PPCF's), $g_{ij}(r)$. In the case of liquid water, these are the H-H, O-H, and O-O PPCF's. Of these, $g_{OH}(r)$ is arguably the most important, containing all of the available direct experimental information on the hydrogen bonding in water.

For this reason, numerous attempts have been made for the determination of the PPCF's of liquid water over the last 50, and especially, over the last 20, years, due to the advent of the technique of isotopic substitution neutron diffraction.[1] All of these attempts (see, e.g., Refs. 2–4) followed the standard route to the PPCF's that starts with three independent measurements, most frequently on three samples of different H/D ratio, followed by the separation of partial structure factors via the inversion of the coefficient matrix, the subtraction of the intramolecular structure factor (or molecular form factor) and finally, the Fourier transformation of the intermolecular partial structure factors to yield PPCF's, which are the desired real-space information. This procedure, known as the ''direct'' separation of PPCF's, seemed to work reasonably well for many molecular liquids. Water, however, is special and problematic in many ways: from the structure determination point of view, most of the difficulties arise from its high H content.

H (and D) is hardly seen by x-ray diffraction, therefore at present it is only the O-O PPCF that can be derived safely via this technique. On the other hand, the huge inelastic incoherent neutron cross section of ^{1}H renders the structurally useful information (coherent scattering) from pure H_2O less than 5% of the (neutron-diffraction) signal, which makes the data analysis nearly prohibitively difficult to perform. Coherent inelastic effects are also considerable, as the masses of H and D are both comparable to the mass of the neutron, which means that a large part of the diffraction signal corresponds to inelastic scattering (dynamic effects). That is, the validity of the assumption that H and D are interchangeable should be confirmed, since H and D are surely not interchangeable dynamically. Therefore, strictly speaking, for studying the PPCF's of liquid H_2O (or D_2O) one would need three independent measurements on H_2O (or D_2O). This can only be achieved by applying also electron diffraction that, although was once used for measuring liquid D_2O,[5] is mostly considered as not sufficiently well suited for investigating condensed phases routinely.

There are a number of problems with the approach itself that has been described above. There are only two ''sure'' (or ''problem free,'' to some extent) experiments aiming at the structure of liquid water: x-ray diffraction on H_2O (or D_2O) and neutron diffraction on D_2O. For water, which is a two component sample, three independent measurements are necessary to obtain PPCF's via the approach described above. Therefore at least one hydrogeneous sample must be measured by neutrons in any practical combination of measurements. The structure factor of a sample with low H content differs only little from the diffraction pattern of pure D_2O (that is, the conditioning of the separation matrix will be poor), while increasing the H content increases the level of systematic errors rapidly. During direct separation of the partials there is no possibility for weighting the different measurements according to their reliability: a neutron-diffraction measurement with a high probability of systematic errors (on H_2O) is treated equally to a measurement with almost no systematic errors at all (D_2O). The most worrying, although not very much publicized, consequence of these problems with liquid water is that PPCF's derived by direct separation seem to tend to be the most consistent with the worst measurement (on H_2O) and the least consistent with the best one (on D_2O) (see Figs. 4 and 7 of Ref. 3). (This is an empirical observation, valid only for the cases inspected here.) The situation clearly needs improving and this is precisely what is attempted here.

First of all, a method is needed that could characterize/quantify the reliability of a given measurement. For this purpose, reverse Monte Carlo (RMC) modeling[6] was shown to be adequate.[7] RMC is a Monte Carlo based tool for generating structural models that are consistent with a given (set of) diffraction data. In principle RMC can access any particle arrangement (i.e., any point of the configuration space) that is consistent with a (set of) physically meaningful constraint(s) imposed on the configuration space. A diffraction data set is precisely such a constraint; the definition of the molecular shape is another example. Therefore if RMC cannot produce a model whose structure factor fits a measured structure factor within errors then the measured structure factor and/or the way it is analyzed should be considered as physically meaningless. Since in general there may be more apparent reasons why a finite model cannot fit a given set of data, let us constrain ourselves to the case of water here. If the constraint of the molecular shape is not applied (that is, there are only free atoms, eventually of very small size, if desired, in the system) then the above statement clearly holds in a reasonably large system. Also, if—as is the case with liquid water, see below—a total structure factor (TSF) of a given isotopic composition, e.g., D_2O, can be modeled without difficulty then it is very hard to explain otherwise why the TSF of another composition, e.g., H_2O, proves to be impossible to model. Naturally, being able to fit some diffraction data within errors does not automatically mean that the data in question are reliable, rather than that the data may be reliable.

Second, measurements from different (neutron, x-ray, electron) sources have not been routinely combined (with the notable exception of Ref. 8; unfortunately, the electron-diffraction TSF, Ref. 5, seems to have similar problems as shown in Fig. 2), even though neutrons are not very sensitive to the O-O pairs and x rays are not very sensitive to H (or D). Clearly, the advantages of all techniques ought to be combined and exploited; this can be very easily done by using the RMC method.

Third, RMC is capable of generating partials from a set containing any number of measurements (for water, this number can be between one and the total number of experiments carried out so far). The reliability of the partials obtained may be radically different depending on the number and also, on the quality of these data sets. During the RMC process, all the experimental data sets are considered as constraints on the available configuration space and modeled simultaneously.

The proposed approach to the partial pair correlation functions of water then:

(i) model (by RMC) all available single structure factors, one by one, to select the ones that may be reliable;

(ii) first combine the two most reliable ones only (x rays on H_2O/D_2O and neutrons on D_2O) and calculate the PPCF's directly from the coordinates;

(iii) add more data sets (as many as possible) and calculate PPCF's directly;

(iv) add constraints to explore the range where the solutions are valid (and ''unique'').

During the reverse Monte Carlo calculations described here, an identical simulation box of 2000 molecules was used, with a molecular density of 0.033 $Å^{-3}$. The flexible molecules were defined as sets of atoms that were kept together via a simple form of coordination constraints.[10] The intramolecular O-H and H-H distances were allowed to vary between 0.9 and 1.1 Å and 1.5 and 1.65 Å, respectively. At least 1.2 million moves (roughly 200 accepted moves/atom) were completed in each calculation. Where convergence could be achieved at all, about 30% of the moves were sufficient to reach satisfactory agreement between measured and model total structure factors.

The first step was to model, one by one, a wide selection of neutron and x-ray total structure factors, in order to find the ones that may be considered for further analysis. For the neutron data, one set of reactor data[3] and one set of pulsed source data[2] were chosen. Additionally, the x-ray structure factor of Narten and Levy[11] was modeled. In general, neutron structure factors of D_2O and the x-ray structure factor of H_2O could easily be reproduced [see Figs. 1(a) and 1(c), respectively]. The quality of the fit to the neutron structure factor of DHO (with 67% D and 33% ^1H) (Ref. 3) was also acceptable [Fig. 1(b)]. On the other hand, none of the neutron H_2O structure factors could be approached; the situation with even the most recent pulsed source data[2] is truly appalling (Fig. 2). It is obvious that if these H_2O neutron data are combined with more reliable data then the quality of the outcome, the set of PPCF's, is rather dubious. (A detailed account of all the calculations will be published later.[9])

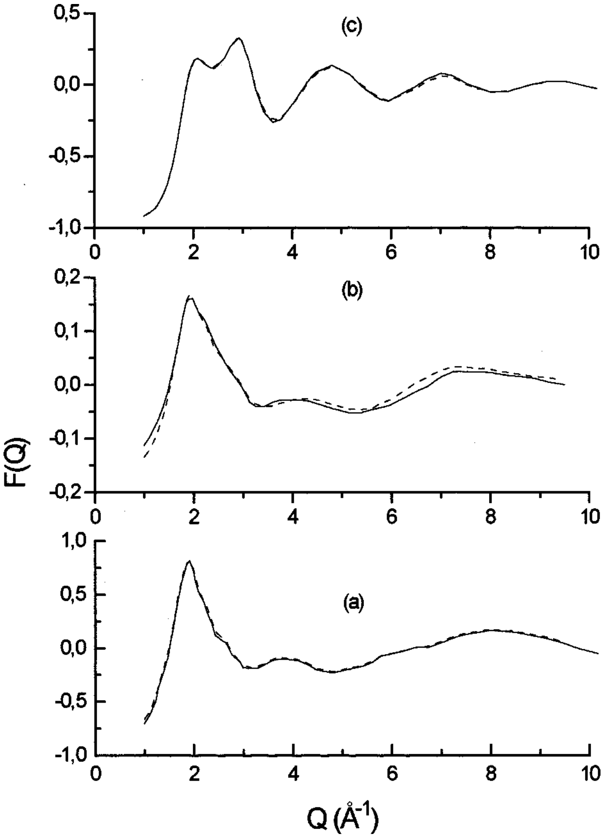

FIG. 1. Experimental (solid) and RMC fitted (dashed) structure factors for liquid water. (a) D_2O (neutron) from Ref. 2; (b) DHO (neutron) from Ref. 3; (c) H_2O (x rays) from Ref. 10. Note that these curves have been obtained when the three TSF's were modeled together; RMC fits to the individual experimental TSF's were slightly better.

The next step was to model the two most reliable data sets, that is, the x-ray TSF of H_2O and the neutron TSF of D_2O. The quality of the fits to the experimental TSF's did not change when these two data were used simultaneously, as shown by Figs. 1(a) and 1(c). The corresponding set of PPCF's is shown by Fig. 3, compared to the widely accepted set of Soper, Bruni, and Ricci.[2] The most striking difference is seen on the O-H partial: the position of the first intermolecular peak, characteristic to the hydrogen bonding, has shifted from about 1.85 Å up to about 2.2–2.3 Å and become less distinct. This result, which is in complete disagreement with any previous suggestion, may well be due to the incomplete input information (two independent TSF's vs the necessary three) and to the fact that the intramolecular and the first intermolecular H-H and the first intermolecular O-H distances lie very close to each other. Differentiating between them is difficult and needs experimental data of high quality,

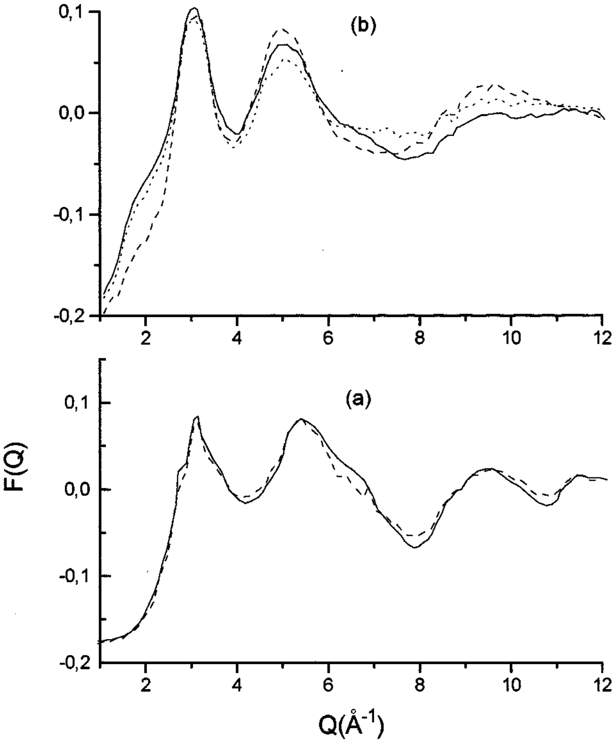

FIG. 2. Experimental (solid) and RMC fitted (dashed) neutron structure factors for H_2O. (a) from Ref. 3; (b) from Ref. 2. In the latter case, a second RMC calculation was also carried out (dotted line) where not even the molecular structure was defined, that is, free atoms were moved; the situation has improved only very little by releasing this important constraint. (Modeling the electron-diffraction results of Ref. 5 gave qualitatively similar results.)

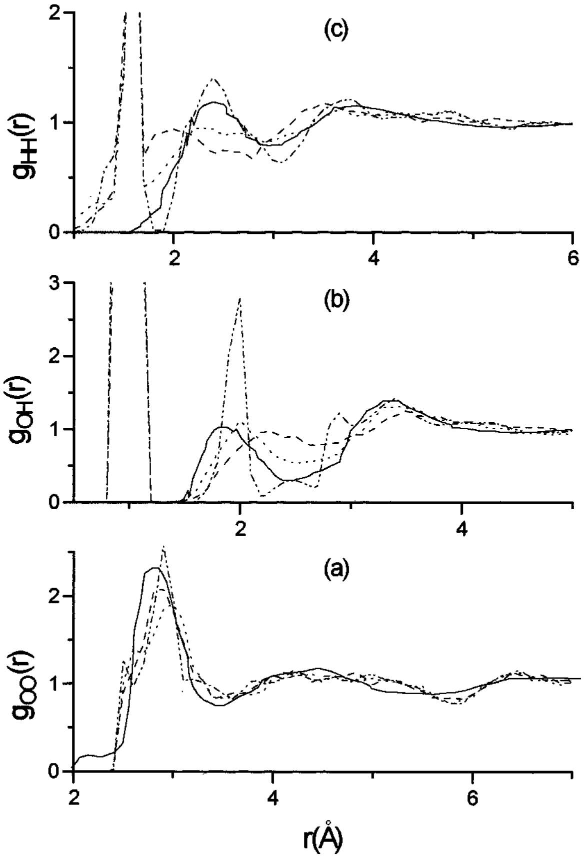

FIG. 3. (a) O-H/D, (b) H/D-H/D, and (c) O-O partial pair correlation functions for liquid water. Solid lines: from Ref. 2; dashes: by fitting the two "sure" TSF's [H_2O with x-rays (Ref. 10) and D_2O with neutrons (Ref. 2)]; dots: DHO with neutrons (Ref. 3) added; dash-dotted: three coordination constraints added (see text). Note the different *r* ranges of the *x* axes.

which was shown to be missing in the case of neutron structure factors of ^1H_2O. It should also be noted that the set of PPCF's derived by RMC on the basis of the two most reliable measurements is consistent with these two experimental TSF's, whereas previously derived sets of PPCF's are not. (Other differences between the new set and the previously derived ones will be discussed in a subsequent paper.[9])

The effects of adding more data sets, taken on H-containing samples, have also been investigated. This way the problem of separation is better determined, although the overall quality of the data applied for the separation is probably not so good as was one step earlier. It was found that if the neutron structure factor of DHO from Ref. 3 was added to the x-ray TSF of H_2O and the neutron TSF of D_2O then all these three TSF's could still be modeled together at a satisfactory level (see Fig. 1). The corresponding partial pair correlation functions are shown in Fig. 3. It is quite clear that the peak on the O-H partial that corresponds to the H bonding shifted to lower *r* values, but still not as much as down to 1.85 Å, as was suggested before. Correspondingly, the position of the first H-H intermolecular peak has shifted towards larger *r* values: the close connection between these two (O-H and H-H) PPCF's, which makes the data analysis difficult, was thus demonstrated again.

During the reverse Monte Carlo process, it is possible to prescribe the number of neighbors of a specified type around a specified type of atom within specified distances, via coordination constraints. In this case, the final configuration of particles (the structural model) must not only be consistent with the experimental data, but at the same time, with the geometrical constraints, as well. Here, coordination constraints have been applied for the purpose of checking how close the PPCF's of Ref. 2 could be approached while remaining consistent with the experimental structure factors. (That the PPCF's of Ref. 2 cannot be consistent, within errors, with the two most reliable TSF's has been established in an unsuccessful attempt when the PPCF's of Ref. 2 were modeled simultaneously with the x-ray TSF of H_2O and the neutron TSF of D_2O.) In the present calculation, three coordination constraints have been imposed, aiming at shifting the first O-H intermolecular peak towards lower *r* values and making it as distinct as possible. The main tool for this was requiring exactly one neighboring O atom around each H atom, between 1.7 and 2.0 Å. This requirement could be fulfilled to the extent of more than 90%, without seriously degrading the level of agreement with the three TSF's (see Fig. 1). The corresponding PPCF's are shown in Fig. 3. As is evident, the position of the O-H peak characterizing H bonding has to lie as close to 2.0 Å, within the constraints applied, and cannot be at a position less than 1.95 Å. This peak can be made rather sharp, as shown in Fig. 3, but in this case, additional small maxima appear between 2.2 and 3.0 Å (and, also, on the O-O PPCF), whose significance is questionable. Physically, this larger O-H intermolecular distance would mean that the O-H$\cdots$O "bond angle" should be less than

180°.

It is rather hard to select the set of PPCF's that could be considered as the ''best'' one. We tend to suggest that at this level, the partials derived from three measurements without using additional constraints can be accepted the most easily. Still, we do not know the ''one and only'' set of partial pair correlation functions of even ambient liquid water (under near critical, or supercritical conditions, the situation is worse, due to the more difficult experiment). In this paper, many sets of PPCF's of liquid water were introduced, all of which are more self-consistent and moreover, more consistent with experimental (diffraction) data than any of the previously suggested sets of partials. The approach suggested here is applicable for exploring the range of possible solutions, which is clearly an advantage over the traditional way. However, since these sets, and in particular, the various O-H and H-H partial $g(r)$'s are largely different, while being consistent with experimental data at the same level, it is concluded that on the basis of diffraction data it is at the moment not possible to derive a definitive set of PPCF's for liquid water. Considering the issue of hydrogen bonding, it is most important to state clearly that no reliable diffraction data are (neither any can be made) consistent with a first intermolecular O-H distance of less than about 1.95 Å. This does not mean that that particular distance cannot be at 1.8–1.9 Å. On the other hand, it does mean to say that no sensible diffraction data on liquid water can be quoted as the basis of our present day picture of the structure of liquid water. Note that knowing the partial pair correlation functions is not the same, but much less, than knowing ''the structure'' of water. In particular, any further discussion about the H bonding, based on present day diffraction data, should be considered as mere speculation.

Possessing a reliable neutron structure factor of ^1H_2O would possibly help a great deal in distinguishing between the otherwise strongly overlapping O-H and H-H partials, since the first and second O-H maxima would appear as negative ''peaks.'' However, it was shown here that, at present, reliable neutron structure factors of ^1H_2O are nonexistent. The consequences of this fact are rather serious, since hydrogenated samples have been widely used for drawing conclusions on the nature of hydrophobic hydration (see, e.g., Ref. 12), for instance, which is of particular importance in biology. According to the findings of the present work, all the partials derived on the basis of heavily hydrogenated samples and naturally, the conclusions drawn on the basis of these partials, may be of suspect.

There may be ways out of the trouble, within the framework of the procedure suggested above. One could pursue more precise neutron measurements and particularly, more careful data analyses subsequently, on ^1H_2O. Another way may be an attempt to evaluate x-ray data differently, so that at least the O-H contribution to the total structure factor could be considered. A new, reliable electron-diffraction measurement would also help. Yet another possibility may be to find other firm experimental evidence, even very little pieces, that could be incorporated in the RMC modeling as constraints. Nevertheless, the findings of the present work should warn us that even if some, or all, of these attempts were successful, a definitive set of partial pair correlation functions, and therefore the definitive structure, of liquid water may remain hidden for long.

The author wishes to thank Dr. Robert McGreevy of the Studsvik Neutron Research Laboratory (NFL) and Professor John Finney (University College, London) for useful discussions. He is also grateful for the hospitality of the staff while staying at NFL Studsvik. L.P. acknowledges financial support from the Bolyai Scholarship Foundation of Hungarian Academy of Sciences.

---

[1] J. E. Enderby, D. N. North, and P. A. Egelstaff, Philos. Mag. **14**, 961 (1966).

[2] A. K. Soper, F. Bruni, and M. A. Ricci, J. Chem. Phys. **106**, 247 (1997).

[3] W. E. Thiessen and A. H. Narten, J. Chem. Phys. **77**, 2656 (1982).

[4] J. C. Dore, in *Water Science Reviews*, edited by F. Franks (Cambridge University Press, Cambridge, England, 1985), Vol. 1.

[5] E. Kálmán, G. Pálinkás, and P. Kovács, Mol. Phys. **34**, 505 (1977).

[6] R. L. McGreevy and L. Pusztai, Mol. Simul. **1**, 359 (1988).

[7] L. Pusztai, O. Gereben, and A. Baranyai, Phys. Scr. **T57**, 69 (1995).

[8] G. Pálinkás, E. Kálmán, and P. Kovács, Mol. Phys. **34**, 526 (1977).

[9] L. Pusztai (unpublished).

[10] L. Pusztai and R. L. McGreevy, NFL Studsvik Annual Report for 1996, Report No. OTH:21, 1997 (unpublished).

[11] A. H. Narten and H. A. Levy, J. Chem. Phys. **55**, 2263 (1971).

[12] J. Turner and A. K. Soper, J. Chem. Phys. **101**, 6116 (1994).