The structure of dimerizing fluids from "experimental" diffraction data by reverse Monte Carlo modelling

L. Pusztai

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

H. Dominguez and O.A. Pizio Instituto de Quimica de la Universidad Nacional Autónoma de México, Coyoacan 04510, México D.F., México

Recibido el 26 de abril de 2002; aceptado el 13 de diciembre de 2002

Molecular dynamics simulations of systems consisting of monoatomic and diatomic species, which mimick equilibrium distributions of particles in dimerizing associating fluids under given external conditions have been performed. The results of these simulations are given in terms of pair distribution functions of atoms and of the corresponding structure factors. The data are considered as "experimental input" for the reverse Monte Carlo (RMC) modelling focused in the evaluation of the (a priori, unknown) composition of the system, and at getting insight into structural properties that are implicitely contained in the experimental structure factor. We discuss usefulness of the simulation scheme, its accuracy and limitations to intend the application of the RMC modelling for more complex associating fluids.

Keywords: Chemically associating liquids; difraction data; reverse Monte Marlo modelling.

Se realizaron simulaciones de dinámica molecular de sistemas formados por especies momoatómicas y diatómicas, las cuales describen distribuciones de partículas en equilibrio de un fluido con dimerización bajo condiciones externas. Los resultados de estas simulaciones se presentan en forma de distribución de los pares de átomos y también de sus factores de extructura. Estos resultados se toman como los experimentales y se usan como datos de entrada para ser usado en el modelo de Monte Carlo inverso, el cual determina la composición del sistema (desconocida apriori) y también las propiedades estructurales, que están contenidas de manera implícita, en los factores de estructura. Se discute la utilidad del procedimiento de simulación, su exactitud y limitaciones para intentar su aplicación en modelos de fluidos más complejos.

Descriptores: Líquidos químicamente asociativos; datos de difracción; modelo de Monte Carlo inverso.

PACS: 61.25.-f, 61.43.Bn

1. Introduction

Computer simulation methods are powerful tools to obtain the structure and thermodynamics of fluids and fluid mixtures (see, *e.g.* Refs. 1-3). Usual implementation of a computer simulation method involves a model Hamiltonian of the system as an input. Next, follows the application of either stochastic (Monte Carlo) or deterministic (molecular dynamics) algorithm to find the properties of interest under given external conditions.

A comparison of the simulation data with the results of theoretical approaches obtained for the same Hamiltonian model is useful to establish validity and accuracy of the theory and its predictions. Also, it is generally accepted that one of the merits of computer simulations is to guide a researcher in adjusting the form of interparticle interactions and their parameters to put the model as close as desired to real systems. In spite of being formally correct, it is a quite difficult task to realize in practice.

Undoubtedly experimental data remain the primary source of knowledge for the description of the equilibrium properties of complex liquid systems. Then, the main question comes out: is it possible to extract and further, how to utilize, all the information contained (even implicitely) in experimental results?. One of the methods that rest heavily on a particular experimental output, such as diffraction data, and combine these data with simulation techniques is the reverse Monte Carlo (RMC) modelling [4-6]). The molecular dynamics counterpart of the procedure also has been developed recently [7]. It is worth mentioning that in the RMC framework, series of configurations (sets of particle coordinates) are generated randomly (by randomly displacing randomly chosen particles one by one in the simulation box). Some of the configurations are accepted in accordance with their compatibility with the experimental structure factor, S(Q) (or the inverse Fourier transform of the structure factor, the pair distribution function, g(r)). This approach has been applied in several systems such as simple fluids and their mixtures, molten salts and alloys, as well as for hydrogen-bonded liquids [8-11].

Our focus in the present study is the implementation of the RMC approach to one class of fluids that have been left out of attention in previous work in this area. Namely, our interest is to establish how the RMC methodology can be applied to chemically associating, *i.e.* reacting, fluids. Furthermore, it would be important to establish if there is a possibility to extract detailed data for the microscopic structure that are contained in the total structure factor of such fluids. We would also like to evaluate how precisely can be the equilibrium concentration of species in the chemically reacting fluids, at given conditions, derived on the basis of the experimental structure factor.

In spite of several successful applications of diffraction techniques to simple fluids, chemically reacting fluids have been rarely examined by diffraction. On the other hand, theoretical research in chemically reacting fluids has flourished during the last decade, benefitting in several aspects from the development of Wertheim's theory of such fluids [12]. This theory had stimulated computer simulations for specific model Hamiltonians (for dimerizing, polymerizing and network-forming fluids (see Refs. 13,14)). The simulations have been focused both on structural properties and on the thermodynamics of association. On the other hand, Smith and Triska [15] have developed the so-called reactive ensemble Monte Carlo approach which focuses on the description of equilibrium association constants. This method has been used for different systems [16,17]. We decided to benefit from data obtained by the reactive Monte Carlo approach and use those as a part of the 'experimental' input for the model described in what follows. However, the reactive ensemble MC approach does not provide the structural properties of the system. Therefore we have undertaken molecular dynamics simulation in this study, for generating the pair distribution function and from that, the structure factor of the model. The present MD simulation also would be helpful in a wider context. Namely, we hope to apply the reverse MD procedure in future work to reproduce the structural properties of the systems in question and get insight into dynamical properties as well.

To summarize, we have reliable data for the equilibrium association constants from the reactive ensemble MC approach [18] and precise results for the structural properties that form 'experimental' input for the reverse Monte Carlo simulation. Similar idea, to implement molecular dynamics simulation together with the reverse MC method, recently has been used by Gubbins *et al.* in Ref. 19.

2. Generation of the MD 'experimental' data and RMC calculation

Our interest in this work, is in the simple chemically reacting fluid in which the monomeric species form dimers, $A + A \rightleftharpoons A_2$, as a result of interparticle interactions. such reaction can be observed in nature, see, *e.g.* the dimerization of nitrogen monoxide, $NO + NO \rightleftharpoons (NO)_2$, or possibly, dimerization in carboxylic acids. In this case, the presence of dimers has been confirmed by spectroscopic data [20]. The reaction takes place at certain temperature and pressure and the concentrations of monomers and dimers are related via the equilibrium association constant.

Few years ago a reactive ensemble Monte Carlo simulation of a system describing equilibrium distribution of species in the chemical reaction of dimerization was performed by Gubbins *et al.* [18]. As a result, a set of data relating density of the system with precise values of concentrations of each species under certain temperature is available. These simulations have been performed for gas-like systems, as well as fluids at liquid-like densities. At low density, the prevailing structure of the fluid is determined by intramolecular correlations of atoms, whereas the intermolecular correlations between diatomics, as well as between diatomics and monomers and between monomers are weak. Nevertheless, having in mind that the system in question is of interest for experimental research we would like to consider it as a pseudoexperimental sample. Usually, the experimental output is given as diffraction data concerning the microscopic structure of the system. To provide such 'experimental' data we have performed MD simulations to obtain the structure factor of particles coming from the pair distribution functions. In this manner we have at our disposal exact results both for the composition of the system at several thermodynamic parameters, as well as the structural properties in the direct space and in the Fourier space. All these sets serve as an "experimental" guide and severe tests to apply the reverse MC procedure, which is aimed at the evaluation of detailed features of the local structure, of the composition, by using *solely* the structure factor. All other properties are considered as unknown in this procedure and will be compared with the outputs from the RMC at the final stage of investigation.

At this very first attempt, we wished to answer the following question: how precisely can equilibrium concentrations be derived from the experimental structure factor?

Our strategy for this investigation is the following: first, (equilibrium) molecular dynamics simulations have been carried out for mixtures of monomers and dimers, A and A_2 . Concentrations of species correspond to the equilibrium thermodynamic state of a dimerization reaction. In order to mimick experimental diffraction data we have calculated the structural properties, such as the atom-atom pair distribution function (pdf) by using MD simulation, and then, numerically, its structure factor. The primary objective of these RMC simulations was to find out to what extent diffraction data can quantitatively characterize the simplest chemical equilibrium. In other words, the question is how well can the mole fraction of the monomers (dimers) be retrieved on the basis of particle configurations in case where no extra information was used to the 'diffraction' data.

Out of the parameter sets reported in Ref. 18 the one which resulted in the highest molar fraction of monomers, X_A =0.401, at low density of atoms, $\rho = 3.05 \times 10^{-3} \text{\AA}^{-3}$, was chosen for our first calculation. The main reason for this choice is that we would like to deal with the favourable situation from the point of the diffraction data.

Molecular dynamics (MD) simulations were run in the canonical (N,V,T) ensemble, using the DL-POLY package [21]. The simulated system contained 400 monomers and 600 dimers. The density of atoms 3.05×10^{-3} Å⁻³ results in a box size of approximately 80 Å. The site-site Lennard-Jones parameters are $\sigma = 3.743$ Å and $\epsilon/k_B = 149.1$ K (the cut-off distance for the potential has been chosen equal to the half of the box length). The atoms in a diatomic molecule were

kept at a bondlength by using the SHAKE algorithm [22]. The bondlength equals to the LJ σ -parameter. The temperature for all the simulations was fixed at T=298.2 K. The atomatom pair distribution functions (independent of participation of atoms in either monomers or dimers) were calculated up to the half of the boxlength and then, structure factors were obtained by Fourier transforming them up to $Q_{max}=20$ Å⁻¹.

The main goal of an RMC calculation is to build threedimensional structural models of a system that are consistent (within certain error) with diffraction data. During the RMC procedure, each of the particle in the box is moved according to a Monte Carlo type algorithm, accepting configurations that provide decreasing difference between measured (experimental, or 'pseudoexperimental' in the present study) and calculated (from the particles coordinates) structure factors, until it reaches minimum, and afterwards oscillates around the minimum value. A fraction of configurations which deviate from the experimental data also are included in the course of a run. A set of (representative) configurations then can be exploited for the calculation of various structural and other properties. In the present case, coordination numbers will be reported: their distribution provides description of the concentrations of different associates. Most frequently the RMC is driven by the structure factor, S(Q). However, one can idealize the picture by assuming that 'perfect' data are available and use the pair distribution function, g(r), to drive the RMC procedure. This possibility will be explored in the present study, besides the procedure utilizing Fourier-transform as an input.

The RMC calculations were conducted on systems that were identical in size and density to those used in the MD simulations. However, unlike in the MD case, only atoms were explicitely placed in the RMC simulation box and the "formation" of diatomic species was allowed in the displacement scheme to permit the system of atoms to accomodate according to the 'diffraction data' (that is, according to the pair correlation functions or structure factors obtained from the MD runs). The total number of monomers, corresponding to X_A and the number of particles in the MD calculations, is N=1600. Out of these, 400, *i.e.* exactly 25 % should be monomers, that is, having no neighbours within the first coordination shell. The 'experimental' value of the average first coordination number therefore is 0.75.

A peculiar feature of RMC of substantial importance from the point of view of the present study is its capability for imposing coordination constraints on the structural models (for more details, see, *e.g.*, Ref. [7]). In the following discussion of the results, if specified, constraints were imposed so that the allowed species restricted to monomers and dimers. Technically, it means that within a pre-specified range, particles were not allowed to have 2 or more neighbours: a move that resulted in such a local coordination was immediately rejected. This proved to be reasonable since common diffraction data contain information only about the *mean* coordination number.

3. Results

Our methodology rests on the structure of the system. Figure 1a, we show the atom-atom pdf (the pdf contains the intramolecular part due to atoms belonging to dimer species) and the corresponding structure factor in Fig. 1b.

It is worth mentioning that the g(r) is dominated by the intramolecular peak and it does not oscillate around unity but has a slope towards it. The intermolecular part otherwise is featureless. This behaviour is typical for low density fluids. The structure factor, S(Q), exhibits a large small angle scattering signal and long lasting decaying oscillations: the latter is the consequence of the only one well-defined characteristic distance (the bondlength).



FIGURE 1a. Pair distribution function of atoms, g(r), for the system with atomic density of $\rho = 3.05 \times 10^{-3} \text{\AA}^{-3}$ and bondlength 3.743 Å. MD data: Solid line; RMC result: symbols.



FIGURE 1b. Structure factor, S(Q) for the same system as defined in (a). MD data: Solid line; RMC result: symbols.

The RMC results fit the 'experimental' pdf and the structure factor (see Figs. 1a and 1b, respectively). The quality of the agreement between model and 'experiment' is excellent in both cases. However, fitting the pair correlation function technically proved to be more difficult than fitting the structure factor. In the former case, more accepted moves were necessary and the fraction of attempted moves was small (around 1:50 in the worst case) and even then, the accuracy was not as good as for the structure factor fitting. All these problems are due to the low density (which results in a high level of disorder) and just one sharp intramolecular peak.

In order to calculate the concentrations of monomers and dimers from particle configurations obtained by RMC, first coordination number of an atom have to be determined. One simply counts the number of neighbours around each particle within a pre-specified distance, i.e. in the first coordination sphere. The pre-specified distance is taken to be the upper limit of the intramolecular peak of the pdf.

The distribution of the number of neighbours is given in Table I. We know in advance, that in the original MD configurations, only the monomer and dimer species are present. After performing the RMC modelling we observe, that besides monomers, *i.e.* zero number of neighbours, particles with 2, 3 and 4 neighbours in the first coordination sphere, appear with nonvanishing frequency. The mean coordination number, n, is 0.9, which is higher than the value of 0.75, coming from MD data. The reason of this discrepancy is that in the model system, chosen according to Ref. [18] both intraand inter-molecular correlations contribute to the pdf in the first coordination shell, c.f. Fig. 1a. In spite of very good agreement between the pdf coming from the RMC procedure and the experimental input, we have not obtained agreement between coordination numbers at the comparable level of accuracy.

From the above observation and having in mind that the structure factors (experimental and RMC) coincide, *c.f.*

TABLE I. Coordination number distributions for the model with bondlength 3.743 Å. The C(n) values in the table are the fractions of atoms with n neighbours in the first coordination shell restricted by upper bound of the "intramolecular" part of the distribution function, whereas \bar{n} is the average cordination number.

$\varrho, \mathrm{\AA}^{-3}$	0.0030	0.0030	0.0030	0.0030
fitting	g(r)	S(Q)	g(r)	S(Q)
constrained	no	no	yes	yes
	C(n)	C(n)	C(n)	C(n)
n=0	0.36	0.49	0.46	0.39
n=1	0.42	0.33	0.56	0.61
n=2	0.18	0.13	0.00	0.00
n=3	0.03	0.03	0.00	0.00
n=4	0.01	0.01	0.00	0.00
\bar{n}	0.91	0.73	0.54	0.61

Fig. 1b, we intended to retrieve the coordination numbers from the RMC configurations yielding S(Q), rather than those yielding g(r). The relevant data are given in the second column of Table I. The mean coordination number improves substantially and agrees well with the MD value. Nevertheless, the values for the monomer and dimer concentrations differ from the MD data and from previous, g(r)-biased modelling.

Next, we have carried out RMC calculations of the pdf and the structure factor for the same input data, but with the inclusion of coordination (geometrical) constraint. The essence of the constraint is that no more than one neighbor for any particle is permitted in the first coordination shell, *i.e.* the configurations of particles that do not satisfy this requirement are rejected. The pdf and the S(Q) from the RMC coincide with experimental data practically at the same level like shown in Fig. 1a and therefore, they are not presented. In addition, again, we focused on the distribution of coordination numbers and on the mean coordination number, according to g(r)-biased and S(Q)-biased modelling. Results for coordination numbers are given in the third and fourth columns of Table I. In both cases we observe that the mean coordination number differs significantly from the predictions of the MD simulation. Also, trends of the behavior of the fraction of monomers and dimers from the S(Q)-biased modelling are favorable. In contrast, these fractions from the q(r)-biased modelling are unsatisfactory. It seems that this latter result is the consequence of impossibility of reaching reasonable statistics for monomers in a very dilute system in question under severe geometric constraints applied and very demanding shape of the input data. An addititonal difficulty is due to the overlap of the intra- and inter-molecular correlations inside the first coordination shell which constitutes our constraint. On the other hand, fitting the S(Q) data is not so demanding even if the geometrical constraints are used.

To have a wider insight into the properties of similar systems, we have included into our consideration other models, at twice and four times higher densities and other values of the bondlength parameter, in order to find the range where the concentration of monoatomic species can be retrieved from the RMC approach with the highest accuracy. For this reason, MD simulations of these models with bondlengths 3.0 and 3.2 Å have been carried out. Having in mind our discussion above, we restricted ourselves to S(Q)-biased modelling and focused on the distribution of the coordination numbers. The relevant set of data are given in Tables II and III. We observe, that for the denser systems in question and under the conditions that the intra-molecular correlations are "better" separated in space from the inter-molecular ones, in comparison to our first model, the agreement of the mean coordination number coming from RMC modelling (without and with geometrical constraints) with the MD result (always 0.75) is satisfactory and in some cases, very good. Moreover, better agreement between the RMC modelling and the MD result can be obtained for the fraction of monomers and of dimers, if the procedure is performed by using coordination constraints.

Also, in all the cases studied, the agreement of the structure factors from two (MD and RMC) methods was excellent. In particular, in the model system with the bondlength 3.0 Å and at density 0.0061 Å⁻³ the 'experimental' and RMC structure factors are compared in Fig. 2 (the constrained and unconstrained calculations give equally good agreement).

We employ the configurations used to derive the S(Q) in Fig. 2, for getting insight and discussing the structural properties that come as a by-product from the MD simulation. Namely, we would like to use the MD obtained monomermonomer distribution function, $g_{mm}(r)$, monomer-atom (belonging to a dimer) $g_{ma}(r)$, and atom-atom distribution function, $g_{aa}(r)$, between atoms belonging to different dimers to compare them with corresponding results from RMC modelling.

In our example, 30 independent configurations from the RMC run, contributing to the structure factor shown in Fig. 2, have been collected and the corresponding g(r)'s have been evaluated. A comparison of these distribution functions is shown in Fig. 3. As it can be seen, the $g_{mm}(r)$ agrees reasonable well, whereas disagreement is found for the two other functions, specifically for the height of the first maximum.

TABLE II. Coordination number distributions for the model with bondlength 3.2 Å. The meaning of the symbols is like in Table I.

$\varrho, \mathrm{\AA}^{-3}$	0.0061	0.0061	0.0122	0.0122
fitting	S(Q)	S(Q)	S(Q)	S(Q)
constrained	no	yes	no	yes
	C(n)	C(n)	C(n)	C(n)
n=0	0.46	0.32	0.44	0.31
n=1	0.37	0.68	0.40	0.69
n=2	0.13	0.00	0.14	0.00
n=3	0.03	0.00	0.02	0.00
n=4	0.01	0.00	0.00	0.00
\bar{n}	0.75	0.68	0.75	0.69
-				

TABLE III. Coordination number distributions for the model with bondlength 3.0 Å. The meaning of the symbols is like in Table I

$\varrho, \mathrm{\AA}^{-3}$	0.0061	0.0061	0.0122	0.0122
fitting	S(Q)	S(Q)	S(Q)	S(Q)
constrained	no	yes	no	yes
	C(n)	C(n)	C(n)	C(n)
n=0	0.48	0.29	0.42	0.29
n=1	0.33	0.71	0.42	0.71
n=2	0.14	0.00	0.14	0.00
n=3	0.04	0.00	0.02	0.00
n=4	0.01	0.00	0.00	0.00
\bar{n}	0.76	0.71	0.75	0.71



FIGURE 2. Structure factor, S(Q) for the system with atomic density of $\rho = 6.1 \times 10^{-3} \text{\AA}^{-3}$ and bondlength 3.0 Å. MD data: Solid line; RMC: result symbols.

The situation is similar to the case when three partial pair distribution functions are to be extracted from the only one total structure factor. Taking this into account, the results presented in Figs. 3a-3c, in our opinion, are reasonable.

It has been shown that under certain conditions, reasonable estimates for the equilibrium concentrations for an $A + A \rightleftharpoons A_2$ dimerization reaction can be obtained from diffraction data, by using the Reverse Monte Carlo technique.



FIGURE 3a. Monomer-monomer pair distribution function, $g_{mm}(r)$ for the system with atomic density of $\rho = 6.1 \times 10^{-3} \text{\AA}^{-3}$ and bondlength 3.0 Å. The structure factor was modelled by RMC and coordination constraints were appplied. MD data: Solid line; RMC result: symbols.



FIGURE 3b. Monomer-atom pair distribution function (where the atom belongs to a dimer), $g_{ma}(r)$ for the system defined in (a). MD data: Solid line; RMC result: symbols.

The most successful version of the RMC modelling seems to be the one by using the structure factor. The mean coordination number, as well the distribution the coordination numbers, are reproduced sufficiently well in this scheme. However, it is not envisaged that diffraction experiments would be applied for the determination of the equilibrium constant - for this purpose, more accurate methods are available. The principal utility of the RMC is to provide the pair distribution function reproducing experimental data with high accuracy. It has been shown that fitting the structure factor produces the atom-atom pair distribution functions of good quality, with respect to the MD results. It is worth mentioning that if the equilibrium constant is known then it can be applied as a constraint for structural modelling of equilibrium chemically reacting fluids. In future, we plan to investigate the dimerizing



FIGURE 3c. Atom-atom pair distribution function (where the atoms belong to different dimers), $g_{aa}(r)$ for the system defined in (a). MD data: Solid line; RMC result: symbols.

model of this work in a much wider range of densities and to cover the entire concentration range. Then it would be of interest to apply the method for chain-forming fluids of practical importance. Also, we expect that our results would stimulate the research of associating fluids by diffraction methods.

Acknowledgment

This work was supported by the Hungarian Academy of Sciences and the CONACyT of Mexico, via the bilateral project E130.377/002. H.D. and O.P. acknowledge support under grants CONACyT (Mexico) 37323E and DGAPA (UNAM, Mexico) IN113201. Financial support for L.P. by OTKA -Hungarian Basic Research Fund (Grant No. T 32308) is also acknowledged.

- 1. M.P. Allen and D.J. Tildesley, *Computer Simulation of Liquids* (Clarendon Press, Oxford, 1987)
- 2. B. Smit and D. Frenkel, *Understanding Molecular Simulations* (Academic Press, New York, 1996)
- K. Binder and D.W. Heermann, Monte Carlo Simulation in Statistical Physics Solid-State Sciences 80 (Springer, Berlin, 1988)
- 4. R.L. McGreevy and L. Pusztai, Molec. Simul. 1 (1988) 359.
- 5. R.L. McGreevy, J. Phys.: Cond. Matt. 13 (2001) R877.
- 6. L. Pusztai, J. Non-Cryst. Sol. 227-230 (1998) 88.
- 7. G. Tóth and A. Baranyai, J. Chem. Phys. 114 (2001) 2027.
- M.A. Howe, R.L. McGreevy, L. Pusztai, and I. Borzsák, *Phys. Chem. Liq.* 25 (1993) 205.
- R.L. McGreevy and L. Pusztai, *Proc. Roy. Soc.(London)* A 430 (1990) 241.

- 10. L. Pusztai, Phys. Rev. B 60 (1999) 11851.
- P. Jedlovszky, I. Bakó, and G. Pálinkás, *Chem. Phys. Letts.* 221 (1994) 183.
- 12. M.S. Wertheim, J. Stat. Phys. 35 (1984) 19; ibid. 35 (1984) 35.
- 13. W.G. Chapman, J. Chem. Phys. 93 (1990) 4299.
- Yu.V. Kalyuzhnyi, G. Stell, M.L. Llano-Restrepo, W.G. Chapman, M.F. Holovko, J. Chem. Phys. 101 (1994) 7939.
- 15. W.R. Smith and B. Triska, J. Chem. Phys. 100 (1994) 3019.
- M. Borowko, A. Patrykiejew, S. Sokolowski, R. Zagorski, O. Pizio, *Czech. J. Phys.* 48 (1998) 371.
- 17. M. Lisal, I. Nezbeda, W.R. Smith, J. Chem. Phys. 110 (1999) 8597.

- J.K. Johnsson, A.Z. Panagiotopoulos and K.E. Gubbins, *Mol. Phys.* 81 (1994) 717.
- 19. K.T. Thomson and K.E. Gubbins, *Langmuir* **16** (2000) 5761.
- 20. C. Colominas, J. Teixido, J. Cemeli, F.J. Luque, M. Orozco, J. *Phys. Chem. B* **102** (1998) 2269.
- 21. T.R. Forester and W. Smith, *DL-POLY Package of Molecular Simulations*; *CCLRC*, (Daresbury Laboratory, Daresbury, Warrington, UK)
- 22. J.P. Ryckaert, G. Ciccotti, H.J.C and Berendsen, J. Computational Physics 23 (1977) 327.