

## Interactions from diffraction data: historical and comprehensive overview of simulation assisted methods

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2007 J. Phys.: Condens. Matter 19 335220

(<http://iopscience.iop.org/0953-8984/19/33/335220>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 129.252.86.83

The article was downloaded on 29/05/2010 at 04:27

Please note that [terms and conditions apply](#).

# Interactions from diffraction data: historical and comprehensive overview of simulation assisted methods

**Gergely Tóth**

Institute of Chemistry, Eötvös University, H-1518 Budapest, PO Box 32, Budapest, Hungary

E-mail: [toth@chem.elte.hu](mailto:toth@chem.elte.hu)

Received 7 March 2007

Published 4 July 2007

Online at [stacks.iop.org/JPhysCM/19/335220](http://stacks.iop.org/JPhysCM/19/335220)

## Abstract

A large part of statistical mechanics is concerned with the determination of condensed matter structure on the basis of known microscopic interactions. An increasing emphasis has been put on the opposite situation in the last decades as well, where structural data, e.g. pair-distance statistics, are known from diffraction experiments, and one looks for the corresponding interaction functions. The solution of this inverse problem was searched for within the integral equation theories of condensed matter in the early investigations, but before long computer simulation assisted methods were suggested. The interest in this field showed an increasing trend after some attempts appeared in the late 1980s. Several methods were published in the 1990s, and one–two methods appear annually nowadays.

In this paper a comprehensive and historical overview is given on the solution of the inverse problem with simulation assisted methods. Emphasis is put on the theoretical grounds of the methods, on the choice of possible input structural functions, on the numerically local or global schemes of the potential modifications, on some advantages and limits of the different methods and on the scientific impact of the methods.

## 1. Introduction

Generally, for the calculation of thermo-physical data two attributes of liquids are included in the statistical mechanical equations: the structure of the system and the interaction among the constituents. If one knows both of them, the calculation recipes are usually straightforward. It was suspected many years ago that there should be some relation between the structure and the potential. The knowledge of one of these attributes should be enough to describe the systems entirely. The first approximations started from the interactions. Integral equation theories were developed, and many features of liquids were determined using them. This field cannot be termed as a new one; for example, the excellent book by Hill was published 50 years ago [1]. The computational modelling or simulation dates back to these years, as well. The two basic methods, Monte Carlo simulation [2] and molecular dynamics [3], were published in 1953 and

1958. Both theoretical and computational methods provided results on the structure for known interactions. The theoretical methods were pushed into the background by the more or less exact computational ones in the last decades. In spite of this, there are many systems in which the finite computational capacity hinders the simulations, and one can get only theoretical data, e.g. on colloid and biological systems.

The investigations in the other direction, in the determination of the interactions for known structures, were started a few years later. The initiative work of Johnson, Hutchinson and March was announced in 1963 [4] and was fully published in 1964 [5]. They calculated the interaction potentials of several metals by the use of the Born–Green hierarchy with the Kirkwood superposition approximation and by the use of the Percus–Yevick theory. Their method was followed by many applications on different systems. The integral equation theories were varied (e.g. hypernetted chain, modified hypernetted chain, etc), the starting data were changed (real space and inverse space functions), and also the numerical methods were different. We do not review these theoretical approaches, because our paper concerns simulation assisted methods. Details of the applications of the integral equation theories can be found in the literature, whereof we cite only a probably arbitrary list [4–15].

Up to this point we have used the phrases ‘structure’ and ‘interaction’. Of course, these concepts should be described clearly. The distribution function theories of liquids provides well defined functions for isotropic homogenous liquids (see e.g. [1, 14, 16]). The pair-correlation function,  $g(r)$ , is proportional to the probability of finding a particle at scalar distance  $r$  from another particle with respect to the average density of the system. In multi-component systems, one can define partial functions,  $g_{\alpha\beta}(r)$ s. Here, the central particle is an  $\alpha$ -type one and the other one is a  $\beta$ -type one. The real-space pair-correlation function can be Fourier transformed into inverse space. The inverse-space function is the structure factor,  $S(q)$ . The two transformations are shown in equations (1) and (2).

$$S(q) = 1 + 4\pi\rho \int_0^\infty r^2 [g(r) - 1] \frac{\sin(qr)}{qr} dr \quad (1)$$

$$g(r) = 1 + \frac{1}{2\pi^2\rho} \int_0^\infty q^2 [S(q) - 1] \frac{\sin(qr)}{qr} dq \quad (2)$$

where  $\rho$  means the number density. The structure factor can be determined in diffraction experiments. Theoretically, the real-space and the inverse-space functions contain the same information and the transformation is unique in both directions. Practically, it is not easy to maintain these features between the two functions. In reality the functions are known at discrete values and contain statistical or experimental errors. The transformation of simulated  $g(r)$ s to  $S(q)$ s seems to be less defective nowadays, because the computational sources make it possible to simulate large systems on a sufficiently long timescale. This means that one can get  $g(r)$  functions with correctly decayed long-range part, with small statistical uncertainty [17, 18] and with an adequately fine grid. In contrast, experimental data often contains reasonable error, because one has to remove or correct many factors, e.g. irrelevant diffraction interactions. The situation is more sophisticated under extreme conditions. Furthermore, one measures a linear combined signal of the partial structure factors in molecular and multi-component systems. It is known in calculus that the solution of linear equations is extremely sensitive to any error of the input data. The linear combination of the partials to the total functions is more robust in the modelling case.

In the case of molecular systems, one can define molecular pair-distribution functions, where the orientation of the two molecules appears in the functions. These molecular functions are frequently used for rigid molecules. For flexible molecules and for the cases where an experimental comparison is important, the so-called site–site representation can be used. Here

the meaning of the partial pair-correlation functions is similar to the multi-component case and  $\alpha$  and  $\beta$  indices mean the different atomic types.

One can define higher-order correlation functions as well. Traditionally, the most important higher-order correlation function is the triplet one.  $g(r, s, t)$  means the relative possibility of finding three particles (1, 2 and 3) at the  $r = r_{12}$ ,  $s = r_{13}$  and  $t = r_{23}$  distances with respect to the average density.

The definition of interactions for liquid systems can be found in most of the textbooks [14, 16, 19, 20]. In classical modelling of liquids one can start with a series expansion of the total interaction to one-, two-, three-, etc particle interactions. The calculation of the higher-order terms is expensive. It is usual to stop at the pair, or occasionally at the triplet-particle interactions. However, the neglected terms have large impact. Therefore, the net effect of the higher-order terms is tried to be incorporated into the pair interaction. The resulted pair-interaction function is called effective pair potential ( $u(r)$ ). It differs remarkably from the simple gas-phase two-particle interaction. The simplification limits the number of the real systems that can be modelled with high accuracy. A slight improvement can be achieved with the addition of three-particle interactions, e.g. for covalent systems.

In our paper we would like to present the state of the art of the simulation assisted potential refinement methods, which uses diffraction experimental data. We have limited ourselves to these inputs, since there is a special unique feature of these experimental data. According to our knowledge, no other experimental data have this attribute.

In 1974, Henderson [16, 21] proved that if a system can be described solely by pair-wise additive pair interactions and the system is in equilibrium at a given state point (e.g. temperature and volume), there is a unique and mutual correspondence of the pair-correlation function and the pair potential. The uniqueness is valid up to an arbitrary constant in the pair potential that is set usually to give zero interaction at infinite distance.

$$g(r) \Leftrightarrow u(r) + \text{constant}. \quad (3)$$

The inverse theorem was proved and interpreted several times: by Evans [11, 22], Zwicker and Lovett [23], Baranyai and Tóth [24], and numerically by Jain *et al* [25]. The unique relationship is not limited to pair interactions. It can be extended up to any higher-order distribution and potential functions, if the system can be described entirely by the set of these interactions.

$$\sum_{i=1}^n g_i \Leftrightarrow \sum_{i=1}^n u_i + \text{constant}, \quad (4)$$

where the lower indices denote the order of the interactions and distribution functions. For example, if a system can be described with pair- and triplet-potential functions, the knowledge of the pair- and triplet-correlation functions is enough to determine the interactions (both the pair and the triplet ones). Sometimes it is confused with the statement that, if a system can be described with pair- and triplet-potential functions, the knowledge of solely the pair-correlation functions is enough to determine the interactions (either the pair interaction or both the pair and triplet ones). The latter is not stated in the inverse theorem and it does not seem to be valid. There seems to be a consensus that the information content of pair-distance statistics is not enough to derive any reliable triplet interactions (e.g. [25]). Anyway, one may try to determine higher-order interactions in the knowledge of solely pair-correlation functions, but the resulting triplet interactions should be checked carefully and compared to relevant experimental data.

The inverse theorem is valid for molecular systems as well, if molecular distribution and interaction functions are used. In the case of the site-site representation, the uniqueness is not strictly valid. It is only practically usable, because the site-site distribution functions can be

obtained from the molecular distribution function by integration over some variables, causing a loss of information [14]. In the case of multi-component systems consisting of independent atoms, the inverse theorem is exactly valid in the site–site representation as well.

The uniqueness theorem emphasizes the diffraction data among the other experimental data in the determination of potential functions. Of course, one can find many attempts for other data, e.g. thermo-physical ones. There are plenty of simulation assisted potential-refinement recipes on these other data [26–32], whereof some are analogous to the methods described in our paper (e.g. [32]).

The paper is organized as follows. First we present the different methods in chronological order. Then we group the methods and we detail the main features. The comparison of the different methods is put in a separate section together with a short overview on the impact and the application of the methods.

## 2. Historical overview

In spite of the attempts to solve the inverse problem by theoretical methods and the success of traditional computational modelling in the 1960s and early 1970s, the first simulation assisted method for potential determination was published by Schommers only in 1973 [33]. He proposed a method in which an initial guess of an interaction potential was iteratively refined by comparing the results of simulations at the given  $i$ th variant of the potential to experimental pair-correlation functions. The study was really a pioneering one: only few applications and references were devoted to this method, and it passed almost unnoticed. This period was part of a very fruitful and progressive time for the theoreticians, who used integral equation theories for the solution of the inverse problem. Perhaps a solution with computer simulations seemed to be marginal at that time. Ten years later, Schommers published another paper on the same method [13], in which he compared his method to integral equation theories. Perhaps this paper initiated Levesque, Weiss and Reatto to elaborate a more sophisticated method (LWR method) [34]. They proposed a scheme for the potential update based on the modified hypernetted chain approximation. They claimed that their method was superior to the Schommers method and to popular integral equation methods. They aroused the interest of the scientific community, and they were commented on, as well [35–37]. The number of the applications increased. For example, Dzugasov *et al* proposed a manual scheme in 1988 [38], in which the potential was modified in each iteration by empirical intuition. One year later he applied the LWR scheme and found it feasible [39].

The reverse Monte Carlo (RMC) simulation method was published in 1988 [40]. It came into general use among the experts of diffraction methods. The RMC method did not solve the inverse problem. It stopped somehow at the half way. Using this method one creates three-dimensional configurations of particles, which are more or less consistent with the diffraction data. The method initiated debates and discussions, and turned the scientific interest to the inverse problem. Two theoretical papers appeared at that time on the inverse problem [22, 23] and also the uniqueness of the generated configurations were examined [41].

In 1994 the LWR method was generalized to multi-component systems [42]. In 1995 a new scheme was published by Lyubartsev and Laaksonen [43]. They broke away from the integral equation theoretical background. They proposed a numerical mathematical formula, in which the potential modification is based on the partial derivatives of the pair-correlation function in respect to the potential parameters. Their method is based on the solution of linear equations. It seemed to be less robust than the Schommers and LWR schemes on experimental data. Subsequently Lyubartsev and Laaksonen proposed numerical tricks to overcome the deficiencies [44–46].

As will be shown later, the most widely spread method was published in its original version by Soper in 1996 [17]. It was a reincarnation of the Schommers method generalized to multi-component systems and strengthened with numerical methods to become robust (damping, use of auxiliary potentials). Soper realized the possible advantages of his method over the widely used RMC technique, and made an effort to develop a usable software kit for the routine evaluation of diffraction data with his method. A rather large number of applications were published before long, and the algorithm was modified to be able to use directly on partial structure factors [47] and on total structure factors [48].

A different approach was proposed by Tóth and Baranyai in 1999 [49, 50]. A direct (non-iterative) method was elaborated and tested, in which three-dimensional configurations were created using the RMC method, and the pair- and the triplet-correlation functions were calculated and were used as the inputs of the Born–Green–Yvon equation (see e.g. [1]). Unfortunately, the RMC method samples the configuration space in a distorted manner [24, 49, 51], and the potentials obtained were not entirely correct.

The Lyubartsev–Laaksonen method was generalized by Tóth in 2001 [52] and in 2003 [53]. The essence of the generalization was to rewrite the formulae to be used directly on total  $S(q)$ s and to avoid the solution of linear equations. In the newer version, the method was able to fit other experimental data, as well. Rutledge proposed a different theoretical approach to solve the inverse problem [54]. On the basis of theory of polydisperse semigrand canonical ensemble, he used a Monte Carlo simulation method to derive pair potentials. Despite the new theoretical concept, the final method was very similar to the Schommers one embedded in a semigrand canonical Monte Carlo method. Rutledge replaced the clearly distinguishable iteration steps of the Schommers method by one long Monte Carlo simulation, in which the potential was refined more continuously. In 2003, Almarza and Lomba chose a numerical approach to create a new algorithm [55]. They also applied a simulation with a continuously refined potential scheme, and they proposed a numerically controlled simple form for the potential modification. In the original version, the input data were the partial  $g(r)$ s, but later they proposed a method with partial  $S(q)$  inputs [56]. The simplicity of their methods was a good basis to extend it to triplet interactions, as well [57]. A non-equilibrium Monte Carlo approach was established by Wilding in 2003 [58]. His method based on the study of Rutledge extended with similar simplification, as was proposed by Almarza and Lomba. A generalized Monte Carlo approach was suggested to fit any kind of distribution, which was sensitive to the variation of function that explicitly or implicitly appeared in the acceptance criteria of Monte Carlo simulations.

The last items do not fully belong to the simulation assisted methods of solving the inverse problem. The force-matching method of Izvekov *et al* [59] was intended for a different task. It projected the many-body force field of quantum calculations onto pair interactions. The neural network method of Tóth *et al* [60] belongs to the solutions of the inverse problem, but there is not any simulation during the application of the method. Several thousand  $u(r)$ – $S(q)$  pairs were calculated by molecular dynamics and these function pairs were used to train artificial neural network. The trained network can be used as a black box to get approximate potential functions on the basis of experimental  $S(q)$ s.

### 3. Details of the methods

#### 3.1. Schommers and Soper

Schommers proposed an iterative method [13, 33] to get pair potentials for known experimental pair-correlation functions. An iterative step consisted of a molecular dynamics or Monte Carlo simulation at given parameters of the potential and a modification of the potential according to

a recipe. The recipe was derived from the diagrammatic density expansion of  $g(r)$  (e.g. [14]).

$$g(r) = \exp\left(-\frac{u(r)}{k_B T}\right) \gamma(r) = \exp\left(-\frac{u(r)}{k_B T}\right) \left(1 + \rho a(r) + \frac{\rho^2}{2!} b(r) + \dots\right), \quad (5)$$

where  $k_B$  is the Boltzmann constant,  $T$  is the temperature and the  $a(r)$ ,  $b(r)$  and further coefficients in  $\gamma(r)$  are sums of different cluster integrals. Schommers supposed that if one finds a  $u_0(r)$  potential that is sufficiently close to the real  $u(r)$  potential, then one can approximate  $\gamma_0(r)$  with  $\gamma(r)$ . Of course, this is not perfectly true, so the potential cannot be determined in one step. An iterative method was proposed and the update of the potential at the  $i$ th iteration was

$$u^i(r) = u^{i-1}(r) - k_B T \ln \left[ \frac{g^{\text{exp}}(r)}{g^{i-1}(r)} \right]. \quad (6)$$

The method could be easily generalized to multi-component systems, where the corresponding partial functions appeared in equation (6) [13]. The extension was successfully applied 13 years later by Soper [17]. At first, it was called the empirical potential Monte Carlo method. We refer to this  $g(r)$  based method as EPMC in this review. Several new features were built in the method:

- (a) The tabulated potential was smoothed at each iteration.
- (b) A damping factor was applied to suppress oscillations and numerical noises at large  $r$ .
- (c) The pair potential was divided into two parts: a reference potential that was not changed during the iteration, and the rest, that was refined in each iteration step.

A suitable choice of the reference potential helped to start with a good initial guess, to keep the structure of the molecule for molecular systems, to take into account the long-range interactions and to take care of the excluded volume. Soper showed also that equation (6) is a convergent recipe, if the system could be described with pair-wise additive pair interactions.

Soper and his co-workers modified the scheme in 1999 [47], and it was renamed the empirical potential structure refinement method. We refer to this technique as EPSR1999. The potential was divided into a harmonic part (responsible for intramolecular interactions), a reference part (which contains mostly Lennard-Jones interactions), a repulsive part (to avoid unphysical overlaps) and an empirical potential part. The latter was modified now on the basis of the difference between the experimental and simulated partial structure factors. It was Fourier transformed from the  $q$  space into the  $r$  one and it was smoothed to remove artefacts of experimental uncertainty and truncation. This  $\Delta_{\alpha\beta}^{i-1}(r)$  difference was used to get the new partial empirical potential.

$$u_{\text{EP}\alpha\beta}^i(r) = u_{\text{EP}\alpha\beta}^{i-1}(r) - k_B T \ln \left[ 1 + \frac{\Delta_{\alpha\beta}^{i-1}(r)}{g_{\alpha\beta}^{i-1}(r)} \right]. \quad (7)$$

Recently, new concepts were introduced into the method [48]. The empirical potentials  $u_{\text{EP}\alpha\beta}(r)$  were represented with a series expansion of Poisson functions. The  $m$ th coefficients of the expansion,  $C_{m\alpha\beta}$ s, remained the same, if one Fourier transformed the potential (to  $u_{\text{EP}\alpha\beta}(q)$ ) and the Poisson functions into the inverse space. The difference between the experimental total structure factor and the simulated one at the  $i$ th iteration of the potential was calculated. The difference was expanded in the inverse-space Poisson functions. We denote these coefficients as  $\Delta C_{m\alpha\beta}^i$ , where  $j$  is the sequence number of the measurement with the upper value of  $N_{\text{meas}}$ . They act on the coefficients of the empirical potentials.

$$C_{m\alpha\beta}^{i+1} = C_{m\alpha\beta}^i + \sum_{j=1}^{N_{\text{meas}}} w_{j\alpha\beta}^{-1} \Delta C_{m\alpha\beta}^i. \quad (8)$$

Soper defined a special weight matrix with  $w_{j\alpha\beta}$  elements, where the traditional coefficients were multiplied with a feedback factor ( $f$ ) of the  $[0; 1]$  interval and the matrix was extended with a diagonal part in which  $(1 - f)$  appeared. The feedback factor can be interpreted as one's trust of the experimental data and  $(1 - f)$  as that of the simulation results. The inverse of the matrix was calculated in a special way [48], and it was used both in the potential modification recipe and in the final determination of the partial structure factors.

### 3.2. The LWR scheme

The Levesque, Weis and Reatto (LWR) scheme is also an iterative method with a simulation and a potential modification in each iteration step [34]. The potential modification part is related to the modified hypernetted chain equation. The first guess potential is obtained by this theory applying a criterion to choose the best hard-sphere bridge function. The success of the method is based on the generality theorem of Rosenfeld and Ashcroft [61]. The iterative refinement steps can be interpreted as corrector steps of the first predictor step. One needs the knowledge of the experimental and the simulated  $g(r)$ s and  $S(q)$ s for the potential modification. It uses the direct correlation function,  $c(r)$ , that can be obtained by the Fourier transformation of the corresponding  $\frac{S(q)-1}{S(q)}$  [14].

$$u^i(r) = u^{i-1}(r) + k_B T \left( \ln \left[ \frac{g^{i-1}(r)}{g^{\text{exp}}(r)} \right] + c^{i-1}(r) - c^{\text{exp}}(r) - g^{i-1}(r) + g^{\text{exp}}(r) \right). \quad (9)$$

Unfortunately, there are several Fourier transformations enhancing the uncertainty of the truncation and discreteness effects. The LWR scheme has been applied on partial functions of multi-component systems as well [42].

### 3.3. Partial derivative methods

The following derivation is according to our previous studies [52, 53], in which we presented a generalization of the Lyubartsev–Laaksonen idea [43]. An average of a quantity  $B_j(\Gamma, \underline{P})$ , which is a function of the phase-space coordinates ( $\Gamma$ ) and the parameters of the potential energy ( $\underline{P}$ ), can be calculated in the canonical ensemble as

$$\langle B_j(\Gamma, \underline{P}) \rangle = \frac{\int B_j(\Gamma, \underline{P}) \exp\left(-\frac{1}{k_B T} H(\Gamma, \underline{P})\right) d\Gamma}{\int \exp\left(-\frac{1}{k_B T} H(\Gamma, \underline{P})\right) d\Gamma}, \quad (10)$$

where  $H(\Gamma, \underline{P})$  is the Hamiltonian of the system.  $\langle B_j(\Gamma, \underline{P}) \rangle$  depends on the  $p_i$  elements of  $\underline{P}$ . The crucial point is the calculation of the partial derivatives in these methods. They can be calculated according to fluctuation formulae [16, 32, 43, 52, 53, 62]:

$$\frac{\partial \langle B_j(\Gamma, \underline{P}) \rangle}{\partial p_i} = -\frac{1}{k_B T} \left( \left\langle B_j(\Gamma, \underline{P}) \frac{\partial H(\underline{P})}{\partial p_i} \right\rangle - \langle B_j(\Gamma, \underline{P}) \rangle \left\langle \frac{\partial H(\underline{P})}{\partial p_i} \right\rangle \right) + \left\langle \frac{\partial B_j(\Gamma, \underline{P})}{\partial p_i} \right\rangle. \quad (11)$$

The partial derivatives of the Hamiltonian are rather simple for most of the potentials. If  $B_j(\Gamma, \underline{P}) \equiv B_j(\Gamma)$ , the last term disappears from equation (11).

Applying a Taylor expansion, a set of linear equations can be constructed in which the  $\Delta p_i$ s are the unknowns.

$$\Delta \langle B_j(\Gamma, \underline{P}) \rangle = B_j^{\text{exp}} - \langle B_j(\Gamma, \underline{P}) \rangle^{\text{simulated}} = \sum_i^{np} \frac{\partial \langle B_j(\Gamma, \underline{P}) \rangle}{\partial p_i} \Delta p_i + O(\Delta \underline{P}^2). \quad (12)$$



Each data point of an experimental quantity, like the points of a structure factor, can be used as a separate  $B_j^{\text{exp}}$ , so one has as many independent linear equations as data points. Since the linear dependence of  $B_j(\Gamma, \underline{P})$  on the potential parameters is seldom fulfilled, the correct parameters cannot be obtained in one step, and the technique has to be repeated in an iterative manner: one iteration step contains a simulation and the modification of the potential. If the number of data points is equal to the number of parameters, one can solve the set of linear equations uniquely. Lyubartsev and Laaksonen proposed this case. They used  $g(r)$ s or  $g_{\alpha\beta}(r)$ s as experimental data, and tabulated pair potentials at the same grids as in the pair-correlation function. In this way, they got a definite answer, but it was very sensitive to the experimental and simulation uncertainties as it is general in the definite solution of linear equations.

We proposed using more data points (equations) than parameters. In this overdetermined case, one can use multidimensional linear fits, or more sophisticated methods, like the Gauss–Newton–Marquardt nonlinear parameter fit (e.g. [63]). Here the merit function can be the maximum likelihood criteria incorporating the uncertainty of the experimental data in the form of weights ( $W$ ). The potential modification in one iteration step was the following for the nonlinear fit:

$$\underline{P}^i = \underline{P}^{i-1} + (J^{(i-1)T} W J + \lambda^{i-1} I)^{-1} J^{(i-1)T} W (\underline{B}^{\text{exp}} - \underline{B}^{i-1}), \quad (13)$$

where  $J$  is the Jacobian,  $I$  is the matrix of unity, and  $\lambda$  is the Marquardt parameter. We chose total  $S(q)$  functions as experimental data. In this way, we avoided two doubtful steps: the separation of the partial functions by solution of linear equations and the inverse to real space Fourier transformation. Instead of them we could simply linearly combine the partials to total functions and we were able to transform sufficiently good quality  $g(r)$ s of simulations to  $S(q)$ s.

We mention here that this potential refinement method is not restricted to diffraction data; it can be used also on other experimental data, as was shown by Bourasseau *et al* [32]. Simultaneous use of different data types is possible, as well [53].

### 3.4. Manual iteration

Experts of diffraction data processing usually have ideas about the relationship among potentials and structural functions. For example, it is widely believed that the excluded volume has a very large impact on the structure of dense liquids. The pair-correlation function is hardly sensitive to potential changes, which do not affect the excluded volume of the particles. This means also, that the non-negligible experimental error of the present instrumentations causes large uncertainty in any potential determination from diffraction data. One can find some studies, where the relationship between the structure factor and the pair potential is discussed [64], also by a sophisticated manner with eigenvectors [65]. Dzugutov *et al* applied another method [38]. They used iterative simulations to reproduce experimental data. The potential was modified several times, e.g. by switching off a different part of the pair potential to find the connection between the potential and the structure. Reasonably good agreement was obtained between the experimental and simulation data. In spite of the conclusion in their first paper, later on the LWR scheme was used by them. They found it more feasible than the manual potential refinement [39]. It is worthwhile noting that Dzugutov's opinion did not agree with the previous studies [64, 65] on the empirical connection between the structure factor and the potential [38].

### 3.5. Method of Almarza and Lomba

In 2003, Almarza and Lomba [55] proposed a method (AL2003) to speed up the usual iterative techniques in which one iteration step consisted of an equilibrium simulation and a potential

variation. They suggested the use of one non-equilibrium simulation, in which the potential was changed before the total equilibration of the system. The potential was regularly refined after small periods of the simulation. The potential variation was rather small and was decreased as the  $g(r)$ s converged to the experimental data.

$$\frac{1}{k_B T} u^i(r) = \frac{1}{k_B T} u^{i-1}(r) + \gamma^{i-1} \frac{g^{i-1}(r) - g^{\text{exp}}(r)}{\Delta g(r)} \langle \Delta g(r) \rangle, \quad (14)$$

where  $\gamma^{i-1}$  denotes that the parameter decreased regularly by multiplication with a number less than one,  $\Delta g(r)$  is the supposed experimental and simulation uncertainty, and  $\langle \Delta g(r) \rangle$  is an average over the  $r$  space of the experimental data.

They modified the recipe, fitting it to partial structure factors (AL2004) [56]. A real-space pair potential was defined as a Fourier transform of an inverse-space potential,  $v(q)$ :

$$\frac{1}{k_B T} u(r) = \sum_{k=1}^{N_q} v(q_k) \frac{\sin(q_k r)}{q_k r}, \quad (15)$$

where  $N_q$  denotes the number of inverse-space components. The Fourier representatives were regularly refined in the simulations:

$$v^i(q_k) = v^{i-1}(q_k) + \gamma^{i-1} \frac{S^{i-1}(q_k) - S^{\text{exp}}(q_k)}{\Delta S(q_k)} \langle \Delta S(q) \rangle. \quad (16)$$

$\Delta S(q_k)$  is the supposed experimental and simulation uncertainty, and  $\langle \Delta S(q) \rangle$  is its  $q$ -space averaged value. The method on the pair-correlation function was extended to triplet-correlation functions by Russ *et al* [57]. Here video microscopy was used to determine both the pair- and the triplet-correlation functions in colloidal suspensions and two- and three-body potentials were successfully derived.

### 3.6. Monte Carlo methods in extreme ensembles

Rutledge suggested a Monte Carlo method in the so-called semigrand ensemble in 2001 [54]. He derived the method on the theoretical background of polydisperse particles in the grand-canonical space. The potential of mean force was generalized for the polydisperse chemical potential. This chemical potential appeared in the acceptance criterion of the Monte Carlo process. Therefore, it affected the chosen distribution function that was a pair-correlation function. The potential function (chemical potential here) was refined regularly during the Monte Carlo simulation. The variation formula was logarithmic, similar to the Schommers method.

In 2003 Wilding generalized the method [58]. He introduced similar numerical considerations to speed up the simulations as Almarza and Lomba did. It was suggested to vary the generalized potential function  $\mu(\sigma)$ , e.g. chemical potential, regularly before the system reaches total equilibrium. The form was similar to that of Almarza and Lomba:  $p^{\text{exp}}(\sigma)$  denoted the desired experimental distribution and  $\gamma$  had the same meaning as in equations (14) and (16).

$$\mu^i(\sigma) = \mu^{i-1}(\sigma) + \gamma^{i-1} \frac{p^{i-1}(\sigma) - p^{\text{exp}}(\sigma)}{p^{\text{exp}}(\sigma)}. \quad (17)$$

The methods of Rutledge and Wilding showed the feasibility of extreme ensemble Monte Carlo simulations in potential refinements. If there is an experimentally determined distribution and one is interested in a potential-like quantity that influences this distribution, one may try to construct a Monte Carlo procedure in a specific ensemble, in which this potential function

emerges in the acceptance criterion of the Monte Carlo procedure. It was shown by these authors that both the Schommers-like form with theoretical background (quasi-equilibrium simulations) and the Almarza and Lomba-like practical recipes (non-equilibrium simulations) were useable here, as well.

### 3.7. Born–Green–Yvon hierarchy and the reverse Monte Carlo method

It is an old result of statistical mechanics, that an exact integro-differential equation links together the pair-correlation function, the triplet-correlation function and the pair potential in equilibrium systems, if the system is in equilibrium and it can be described by pair-wise additive pair potentials (see e.g. [1]).

$$-\frac{d}{dr} \left( \ln g(r) + \frac{1}{k_B T} u(r) \right) = \frac{1}{k_B T} \pi \rho \int_0^\infty ds \frac{du(s)}{ds} \int_{|r-s|}^{r+s} dt \frac{t(r^2 + s^2 - t^2)}{r^2} \frac{g(r, s, t)}{g(r)}. \quad (18)$$

If one knows both the pair- and the triplet-correlation functions of a liquid, theoretically the pair potential can be obtained. This feature was shown by us practically in 1999 [49, 50]. We calculated the triplet- and the pair-correlation functions of Metropolis Monte Carlo configurations, and used the data in the numerical solution of the Born–Green–Yvon equation. The results were approximately correct, but to reach the exact solution one should use infinitely small grids with high statistical accuracy in the determination of the correlation functions. We applied the method also on configurations provided by reverse Monte Carlo simulations. Depending on the details of the reverse Monte Carlo simulation, more or less correct pair potentials were obtained. The discrepancy was caused partly by the inappropriate configuration sampling of the phase space, which is a weakness of the reverse Monte Carlo method [24, 49, 51, 66–68]. The performance of this combined method did not allow using it routinely, but it would become a powerful method, if the correct sampling of the phase space can be achieved in a reverse Monte Carlo-type procedure. The most important advantage would be that it is a non-iterative, direct method.

### 3.8. Force matching

As was mentioned earlier, the starting point of the reverse Monte Carlo modelling is a measured structure factor. It provides three-dimensional configurations, but traditional forces and energies do not appear in the scheme. We may say that the RMC method stops half way in the solution of the inverse problem. In contrast, the force matching methods [59, 69] start at the middle of the inverse problem. Here, the input data are configurations together with the classical mechanical forces acting on the particles, and the results are the classical mechanical force functions. The methods were developed to project the force field of quantum mechanical dynamics (e.g. Car–Parrinello simulation) onto classical mechanical force functions. It can be performed on all atoms of the *ab initio* simulations, or only on a part of the atoms to derive coarse-grained potentials [44]. In the latter case the input data can be obtained by classical mechanical simulations, too. The main goals are a possible speed up and enlargement of the simulation by the use of classical force functions and/or the omission of a part of the particles. Ercolessi and Adams applied analytical potentials, while Izvekov *et al* used cubic-spline force functions augmented with electrostatic interactions, in which the partial charges of the atoms were parameters as well. We note here that the potential matching analogue of the force matching scheme has been developed recently [67].

### 3.9. Neural networks

Recently, an unusual resolving of the inverse problem was published by us [60]. The method of artificial neural network, which showed its power in many successful biological applications, was applied to the problem. Neural networks are mathematical models with adjustable parameters. These parameters can be optimized in a way to find a good nonlinear approximate functional relation between two functions, of which one is the input and the other is the output. The optimization of the parameters is called training, and it is performed on known pairs of input–output functions. Here the parameters of the neural network are set to such values, that the network gives the best answers on input data, where one also knows the output data. Thereafter, the trained network can be used as a black box to get answers on real problems, where one knows only the input functions.

In our case, the input functions were structure factors supplemented with particle densities, and the output functions were pair potentials. A few thousand known structure-factor–pair-potential pairs were calculated on one-component Lennard-Jones, Morse and Buckingham systems by molecular dynamic simulations. The network was trained to these data and the performance of the method was tested on further pairs. The results were convincing, in the sense that in most cases realistic pair potentials were obtained, which could be used at least at the level of starting potentials for iterative potential refinement methods.

## 4. Comparison of the methods

### 4.1. Local or global change

In this section, we would like to discuss and compare some common features of the methods. If one tries to sort the methods, one aspect could be the local or global effect of the differences in the structural function on the potential function. For example, in the method of Schommers and the EPMC, solely the difference at a given  $r$  value in the pair-correlation function causes a change in the potential function at the same  $r$  distance. Similarly, in the AL2003 method and in the extended Monte Carlo methods, the effect may be termed as local. There are several structural functions in the LWR scheme. If one discusses the method in the form of equation (9), the scheme is local, because the structural functions affect the potential function at the same distance. In a detailed view, one recognizes that the real-space direct correlation function is calculated on the basis of the structure factor that is calculated as the whole  $r$ -space Fourier transform of the pair-correlation function. This means that the effect of  $c(r)$  can be termed as the global impact of the difference between the experimental and simulated pair-correlation function. Similarly, the EPSR1999 and the AL2004 schemes can be termed as local or global. If one uses the concept of an inverse-space potential, these methods are local. Thinking only in  $r$ -space potential results in a global method, since the difference in the  $S(q)$ s at a given  $q$  value changes the potential at every  $r$  distance. The EPSR2005, the partial derivate, the combined RMC-BGY and the neural network methods are global. The local–global feature of the different methods is summarized, together with some other aspects, in table 1.

We would like to notice another aspect of the iterative methods. Some of the methods originate from an expansion according to the density of the systems. In these cases, iteration could give a perfect answer only in the ideal gas case. The Schommers, EPMC, Rutledge and Lyubartsev–Laaksonen methods are such ones [58, 66].

### 4.2. Equilibrium or non-equilibrium simulation

The simulation and the potential modification are separate parts of one iteration in the most of the schemes. It is a time-consuming approach, but it is necessary especially in cases in

**Table 1.** Selected aspects of the simulation assisted inverse methods.

	Schommers	EPMC	EPSR1999	EPSR2005	LWR	Lyubartsev Laaksonen	Tóth 2001, Tóth 2003	AL2003	AL2004	Rutledge	Wilding	RMC + BGY	Neural network
Local/global	L	L	L/G	G	G	G	G	L	L/G	L	L	G	G
Iterative/non equilibrium/direct	I		I		I	I	I	NE	NE	I	NE	D	D
Input: $g(r)$ or $S(q)$	$g(r)$	$g(r)$	$S(q)$	$S(q)$	$g(r)$	$g(r)$	$g(r)$ or $S(q)$	$g(r)$	$S(q)$	$g(r)$	$g(r)$	$g(r)$ or $S(q)$	$S(q)$
Input: total or partial functions	P	P	P	T or P	P	P	T or P	P	P	P	P	T or P	P
Fourier transformation: $r$ to $q/q$ to $r$ /no	No	No	$r$ to $q$ and $q$ to $r$	$r$ to $q$	$r$ to $q$ and $q$ to $r$	No	$r$ to $q$	No	$r$ to $q$ and $q$ to $r$	No	No	$r$ to $q$	No
Experimental uncertainty incorporated	No		No	Yes	No	No	Yes	Yes	Yes	No	No	Yes	Impl.
Convergence controlled	No		No	Yes	No	Yes	Yes	Yes	Yes	No	Yes	No	No
Smoothing	No		Yes		No	No	Yes	No	No	No	No	No	No
Approximate scientific impact of the original articles	40		150		60	80	10	10	0	5	15	15	0
Approximate number of applications	10 (+40)		40		10	10	1	1	0	1	1	0	0

which the potential modification recipes contain simulated functions which are sensitive to the fluctuations of the simulations, e.g. the calculation of the partial derivatives or direct correlation functions. The sensitivity can be smeared out, if one allows only small changes in the potential modification by applying constraints. This technique was applied in the AL2003, AL2004 and Wilding methods. These recipes can be interpreted as non-equilibrium methods. There are also direct methods, such as the RMC+BGY or neural network techniques.

#### 4.3. Input data and Fourier transformations

The authenticity of experimental functions decreases rapidly with the number of data processing steps. As was mentioned in the introduction, it seems to be preferable to use total structure factors instead of partial pair-correlation functions as experimental input data. Two consecutive mathematical procedures are necessary to get  $g_{\alpha\beta}(r)$ s from the total structure factors in the classical evaluation of diffraction data. If the number of the measurements with different partial weight is higher than the number of the partials, one can solve a set of linear equations to get partial functions. The second procedure is Fourier transformation between the inverse-space and real-space correspondents. The order of the two mathematical procedures can be changed. This feature is summarized in the third and fourth rows of table 1. The methods were sorted according to the possibility of real-space and inverse-space experimental inputs. Some of the methods can be used separately on both inputs. We have emphasized the possibility of applying total functions or only partial ones. We note here that one can use also partial functions in those methods that can treat total functions (RMC+BGY, Tóth, EPSR2005). In the case of the neural network method, the use of total functions has not been implemented yet.

There is another feature related to the inverse and real nature of input functions. The Fourier transformations are continuous ones in equations (1) and (2), but in the reality both the experiments and the simulations provide discrete data. Also their range is limited. Nowadays the simulations provide mathematically reasonable pair-correlation functions, while the experimental data are less accurate and they are rather limited in the  $q$  range. Therefore, the real-space to inverse-space Fourier transformation is preferred. This experience gave important momentum to the fast spreading of the reverse Monte Carlo technique. The fifth row shows the direction of the Fourier transformations in table 1. There is no Fourier transformation in most cases where the pair-correlation functions are the inputs. There are Fourier transformations in both directions for most of the structure factor based methods, except our previous one (Tóth 2001 and Tóth 2003), the RMC+BGY direct method and the newest method of Soper (EPSR2005). Here, there are transformations only in the preferable  $r$  to  $q$  direction.

#### 4.4. Experimental error

The next comparison of the methods concerns the possible built-in treatment of experimental and simulation uncertainties, smoothing and damping methods. The original recipe of Schommers did not include any of these techniques that decreased its robustness (e.g. [34]). In the LWR scheme the simulated  $g(r)$  was extended to large  $r$  by theoretical methods in order to reduce the truncation error of the two (backward and forward) Fourier transformations. Unfortunately, the extension made the method become slightly arbitrary and less robust. Dzугutov found the extension superfluous [39], while Kahl and Kristufek thought it to be necessary. Lyubartsev and Laaksonen used a damping parameter in the potential refinement step to reduce the changes indicated by the partial derivatives. They applied further empirical numerical control in their program package. Soper built in techniques to the Schommers

method in EPMC: the pair-correlation functions were smoothed and an exponential decay factor was applied to reduce the noise at large  $r$  in the potential. In the EPSR1999 technique, the difference between the experimental and simulation structure factors is Fourier transformed into the  $r$  space, and the function is smoothed there. We mention again, that besides the smoothing and long-range decay functions there are auxiliary potentials in the methods of Soper. These potentials help for example in maintaining of the excluded volumes, and therefore suppress some possible effects of the experimental and simulation uncertainty. In the EPSR2005 method a feedback factor was introduced to quantify one's trust in the experimental data. The feedback factor also played the role of a convergence and experimental error controller. Rutledge's extended Monte Carlo method does not include any method to smooth or take into account the experimental method. In our method (Tóth 2003) the experimental uncertainty was included explicitly, the effect of the partial derivatives is smoothed by the method of Marquardt in the Gauss–Newton parameterization [63], and the potentials were smoothed by Golay–Savitzky formulae [52, 53]. The Marquardt method has similar effects on the calculation of inverse matrices as the method used by Soper with the feedback factor in EPSR2005. In the recipes of Almarza and Lomba the experimental uncertainty is included. Furthermore, the potential modification is controlled via a parameter reducing the change of the potential in one iteration step. A similar parameter can be found in the Wilding method. There were no error treatments in the solution of the Born–Green–Yvon equation part of the RMC-BGY scheme, but the experimental error was incorporated into the reverse Monte Carlo part. The uncertainties are not treated explicitly in the neural network method; this method can be understood as a technique in which the errors are taken into account in an average way.

#### 4.5. Scientific impact of the methods

The last rows in the table 1 are intended for the scientific response and the applications of the methods. The numbers are approximate. We checked the literature by the use of the WEB of Science database [70]. The scientific response was approximately counted as the number of citations on the original article or on a later published fundamental study. The number of applications was estimated as the number of the applications of the given method and the number of studies in which the basic concept of the method is used. We did not check if more articles were published on the same investigation, or if a significantly modified technique was applied. Numerical comparisons of previous methods to a new one were counted into the applications, as well. The references are given only if there is an application or enhancement of the method.

The methods of Soper have had the largest scientific impact and the largest number of applications [47, 71–108]. Of course, the publication priority of Schommers implicitly means that the EPMC, and partly the EPSR, applications belong to the Schommers method too. We found a further ten applications [80, 109–118], in which the simple Schommers technique was used without the developments of Soper. However, the dominating method seems to be the EPSR due to the efforts of Alan Soper, who is trying to develop a routine program package for analysing diffraction data in the framework of the inverse problem.

Two other methods obtained large scientific impact, the LWR scheme and the method of Lyubartsev and Laaksonen. The role of the LWR method is unquestionable. This method focused on the inverse problem at the moment when computational investigations started to be widely accepted and the role of the traditional integral equation techniques decreased. The application number of the LWR method [39, 42, 80, 81, 119–122] shows a decreasing trend in recent years. In contrast, the Lyubartsev–Laaksonen method is becoming popular [44–46, 123–131]. It is quite often used in a task that is not the processing of diffraction

data, but the so-called coarse graining of potentials. This coarse graining (or atom unifying) is used in mesoscopic simulations, in which a part of the molecules is treated as a unit with an overall potential function [110, 132]. The approach is very useful in the speed up and enlargement of simulations on biosystems, polymers and other large molecules. The coarse graining of potentials seems to be an intensively investigated field of computation physics, as one can check by a simple search in the current literature. The general procedure is similar to the inverse problem, but detailed structural information is available, because the structural data are collected by all-atom simulations. The main idea is to omit the unimportant atoms, to fuse the important ones into groups, and determine an effective classical potential that correctly reproduces the spatial distribution of these groups. The Lyubartsev–Laaksonen scheme is often used and proposed for this coarse-graining process. Here the use of  $g(r)$ s instead of  $S(q)$ s is not a disadvantage. The other methods (Almarza and Lomba, Tóth, Rutledge and Wilding) are less popular. Their low impact is possibly due to the lack of user-friendly open program codes and because the authors are partly outsiders in the data processing of diffraction data. Of course there are a few applications. The Almarza–Lomba method was applied to determine three-body interactions on colloids [57], and the Wilding method has some connection to the field of polydisperse hard-sphere systems [133]. The Rutledge method initialized a macromolecular conformation determination algorithm [134]. A variant of Tóth’s method is used in the determination of united atom potentials on the basis of thermochemical data [32, 135]. The RMC-BGY approach seems to be only theoretically important up to now, because it was the first exact numerical application and proof of the Born–Green–Yvon equation.

The simple counting of the applications does not reflect why a method was applied. We emphasize here that the purpose of most applications was not the derivation of interaction potentials. Soper’s methods were used in the context of diffraction data evaluation and the derived potentials were not routinely reused in later application. In contrast, the applications related to the coarse graining of potentials (e.g. the Lyubartsev–Laaksonen method and the applications in [32] and [135]) were devoted to providing interaction potentials for mesoscopic simulations.

## 5. Conclusions

The aim of this paper was to give an overview of the simulation assisted potential determination methods which use experimental diffraction data as input. We provided a historical overview for both the theoretical part of the inverse problem and the different simulation assisted methods. The first method is more than 30 years old, but the methodological research accelerated only in the last decade. The algorithmic details were compared and we recognized some main attempts:

- (1) The improvement of the numerical algorithms to be robust on different data.
- (2) The speed up of the iterative methods by non-equilibrium schemes.
- (3) Trials to use as much primary diffraction information as possible, e.g. total structure factors.
- (4) Inclusion of other intuitive or empirical facts in the modelling.
- (5) Development of program packages.

This last feature is connected to the application of the methods, and seems to be essential for the spread of a method. We compared the different methods and we tried to identify the common and the distinct attributes of the methods. The main features were summarized in table 1.

The original purpose of the simulation assisted methods was to provide interaction functions on experimental diffraction data. Taking into account the scientific impact of the



methods, we acknowledged the empirical structure potential refinement method of Alan Soper as the most popular simulation assisted inverse method in the evaluation of diffraction data. At first, this method was an enhanced-level application of the Schommers method, but after several improvements and methodological changes a new technique was established. The success of this method was supported by the close connection to experiments and by the available program codes.

Recently, inverse methods seem to be important in another field of computational chemistry and physics. The number of coarse-grained simulations has increased in recent years, and this simulation demands an increased number of coarse-grained potentials. In these simulations one needs interaction potentials for groups of atoms, for whole molecules, or for entities embedded in a matrix, where the matrix appears only in an implicit way in the interaction functions. The potentials can be obtained by inverse methods on the data of all-atom simulations, as well. Here other attributes of the inverse methods are more important than in the case of the diffraction data processing. The method of Lyubartsev and Laaksonen seems to be a preferred choice in this case.

### Acknowledgment

The author is grateful for the tenure of a Hungarian Research Fund (OTKA T43542).

### References

- [1] Hill T L 1956 *Statistical Mechanics* (New York: McGraw-Hill)
- [2] Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A N and Teller E 1953 *J. Chem. Phys.* **21** 1087
- [3] Alder B J and Wainwright T E 1958 *Proc. Int. Symp. on Statistical Mechanical Theory of Transport Processes Brussels 1956* ed I Prigogine (New York: Wiley) pp 97–131
- [4] Johnson M D and March N H 1963 *Phys. Lett.* **3** 313
- [5] Johnson M D, Hutchinson P and March N H 1964 *Proc. R. Soc. A* **282** 283
- [6] Howells W S and Enderby J E 1972 *J. Phys. C: Solid State Phys.* **5** 1277
- [7] Ailawadi N K, Banerjee P K and Choudry A 1974 *J. Chem. Phys.* **60** 2571
- [8] Kumaravadivel R, Evans R and Greenwood D A 1974 *J. Phys. F: Met. Phys.* **4** 1839
- [9] Gopala Rao R V and Joarder R N 1978 *Phys. Lett. A* **67** 71
- [10] Croxton T L and McQuarrie D A 1979 *J. Phys. Chem.* **83** 1840
- [11] Evans R 1979 *Adv. Phys.* **28** 143
- [12] Dharma-wardana M W C and Aers G C 1983 *Phys. Rev. B* **28** 1701
- [13] Schommers W 1983 *Phys. Rev. A* **28** 3599
- [14] Hansen J P and McDonald I R 1986 *Theory of Simple Liquids* (London: Academic)
- [15] Sen A K and Rajagopalan R 1991 *Phys. Chem. Liq.* **22** 195
- [16] Gray C G and Gubbins K E 1981 *Theory of Molecular Fluids* (New York: Oxford University Press)
- [17] Soper A K 1996 *Chem. Phys.* **202** 295
- [18] Baranyai A and Tóth G 1997 *J. Chem. Phys.* **107** 8575
- [19] Allen M P and Tildesley D J 1987 *Computer Simulation of Liquids* (Oxford: Clarendon)
- [20] Frenkel D and Smit B 1996 *Understanding Molecular Simulation* (London: Academic)
- [21] Henderson R L 1974 *Phys. Lett. A* **49** 197
- [22] Evans R 1990 *Mol. Simul.* **4** 409
- [23] Zwicker J and Lovett R 1990 *J. Chem. Phys.* **93** 6752
- [24] Tóth G and Baranyai A 2000 *Trends Stat. Phys.* **3** 165
- [25] Jain S, Garge S and Kumar S K 2006 *Ind. Eng. Chem. Res.* **45** 5614
- [26] McDonald I R and Singer K 1969 *J. Chem. Phys.* **50** 2308
- [27] Njo S L, van Gunsteren W F and Müller-Plathe F 1995 *J. Chem. Phys.* **102** 6199
- [28] Bembenek S D and Rice B M 1999 *Mol. Phys.* **97** 1085
- [29] Bembenek S D and Rice B M 2000 *J. Chem. Phys.* **113** 2354
- [30] Ungerer P, Beauvais C, Delhomelle J, Boutin A, Rousseau B and Fuchs A H 2000 *J. Chem. Phys.* **112** 5499
- [31] Zerarka A, Boumedjane Y and Hans J 2002 *Phys. Rev. A* **66** 052717

- [32] Bourasseau E, Haboudou M, Boutin A, Fuchs A H and Ungerer P 2003 *J. Chem. Phys.* **118** 3020
- [33] Schommers W 1973 *Phys. Lett. A* **43** 157
- [34] Levesque D, Weis J J and Reatto L 1985 *Phys. Rev. Lett.* **54** 451
- [35] Dharma-wardana M W C and Aers G C 1986 *Phys. Rev. Lett.* **56** 1211
- [36] Levesque D, Weis J J and Reatto L 1986 *Phys. Rev. Lett.* **56** 1212
- [37] Schommers W 1987 *Phys. Rev. Lett.* **58** 427
- [38] Dzugutov M, Larsson K E and Ebbsjö I 1988 *Phys. Rev. A* **38** 3609
- [39] Dzugutov M 1989 *Phys. Rev. A* **40** 5434
- [40] McGreevy R L and Pusztai L 1988 *Mol. Simul.* **1** 359
- [41] Pusztai L and Tóth G 1991 *J. Chem. Phys.* **94** 3042
- [42] Kahl G and Kristufek M 1994 *Phys. Rev. E* **49** 3568
- [43] Lyubartsev A P and Laaksonen A 1995 *Phys. Rev. E* **52** 3730
- [44] Lyubartsev A P and Laaksonen A 1997 *Phys. Rev. E* **55** 5689
- [45] Lyubartsev A P and Laaksonen A 1999 *J. Chem. Phys.* **111** 11207
- [46] Lyubartsev A P and Laaksonen A 2000 *Chem. Phys. Lett.* **325** 15
- [47] Yamaguchi T, Hidaka K and Soper A K 1999 *Mol. Phys.* **96** 1159
- [48] Soper A K 2005 *Phys. Rev. B* **72** 104204
- [49] Tóth G and Baranyai A 1999 *Mol. Phys.* **97** 339
- [50] Tóth G and Baranyai A 2000 *J. Mol. Liq.* **85** 3
- [51] Tóth G and Baranyai A 1997 *J. Chem. Phys.* **107** 7402
- [52] Tóth G 2001 *J. Chem. Phys.* **115** 4770
- [53] Tóth G 2003 *J. Chem. Phys.* **118** 3949
- [54] Rutledge G C 2001 *Phys. Rev. E* **63** 021111
- [55] Almaraz N G and Lomba E 2003 *Phys. Rev. E* **68** 011202
- [56] Almaraz N G, Lomba E and Molina D 2004 *Phys. Rev. E* **70** 021203
- [57] Russ C, Brunner M, Bechinger C and von Grünberg H H 2005 *Europhys. Lett.* **69** 468
- [58] Wilding N B 2003 *J. Chem. Phys.* **119** 12163
- [59] Izvekov S, Parrinello M, Bunham C J and Voth G A 2004 *J. Chem. Phys.* **120** 10896
- [60] Tóth G, Király N and Vrabcz A 2005 *J. Chem. Phys.* **123** 174109
- [61] Rosenfeld Y and Ashcroft N W 1979 *Phys. Rev. A* **20** 1208
- [62] Lagache M, Ungerer P, Boutin A and Fuchs A H 2001 *Phys. Chem. Chem. Phys.* **3** 4333
- [63] Press W H, Teukolsky S A, Vetterling W T and Flannery B P 1992 *Numerical Recipes in Fortran* (Cambridge: Cambridge University Press)
- [64] Jacobs E and Andersen H C 1975 *Chem. Phys.* **10** 73
- [65] Ballentine L E and Jones J C 1973 *Can. J. Phys.* **51** 1831
- [66] Tóth G and Baranyai A 2005 *J. Phys.: Condens. Matter* **17** 159
- [67] Tóth G 2006 in preparation
- [68] Tóth G, Pusztai L and Baranyai A 1999 *J. Chem. Phys.* **111** 5620
- [69] Ercolossi F and Adams J 1994 *Europhys. Lett.* **26** 583
- [70] WEB of Science (<http://isiknowledge.com> accessed at August 2006)
- [71] Soper A K 1996 *Faraday Discuss.* **103** 41
- [72] Bowron D T, Finney J L and Soper A K 1998 *J. Phys. Chem. B* **102** 3551
- [73] Bowron D T, Finney J L and Soper A K 1998 *Mol. Phys.* **93** 531
- [74] Andreani C 1998 *J. Mol. Liq.* **78** 217
- [75] Soper A K 1998 *J. Mol. Liq.* **79** 179
- [76] Yamaguchi T and Soper A K 1999 *J. Chem. Phys.* **110** 3529
- [77] Santoli G, Bruni F, Ricci F P, Ricci M A and Soper A K 1999 *Mol. Phys.* **97** 777
- [78] Yamaguchi T, Benmore C J and Soper A K 2000 *J. Chem. Phys.* **112** 8976
- [79] Koh C A, Wisbey R P, Wu X P, Westacott R E and Soper A K 2000 *J. Chem. Phys.* **113** 6390
- [80] Neufeind J, Fischer H E and Schröder W 2000 *J. Phys.: Condens. Matter* **12** 8765
- [81] Neufeind J, Fischer H E and Schröder W 2000 *Physica B* **276** 481
- [82] Soper A K 2000 *Chem. Phys.* **258** 121
- [83] Bowron D T, Soper A K and Finney J L 2001 *J. Chem. Phys.* **114** 6203
- [84] Landron C, Hennem L, Jenkins T E, Creaves G N, Coutures J P and Soper A K 2001 *Phys. Rev. Lett.* **86** 4839
- [85] Soper A K 2001 *Mol. Phys.* **99** 1503
- [86] Landron C, Soper A K, Jenkins T E, Greaves G N, Hennem L and Coutures J P 2001 *J. Non-Cryst. Solids* **293** 453
- [87] Bowron D T and Moreno S D 2002 *J. Chem. Phys.* **117** 3753
- [88] Dixit S, Soper A K, Finney J L and Crain J 2002 *Europhys. Lett.* **59** 377

- [89] Moreno S D and Bowron D T 2002 *J. Phys. Chem. B* **106** 11876
- [90] Hardacre C, McMath S E J, Nieuwenhuyzen M, Bowron D T and Soper A K 2003 *J. Phys.: Condens. Matter* **15** 159
- [91] Hardacre C, Holbrey J D, McMath S E J, Bowron D T and Soper A K 2003 *J. Chem. Phys.* **118** 273
- [92] Bowron D T and Moreno S D 2003 *J. Phys.: Condens. Matter* **15** 121
- [93] Buchanan P, Soper A K, Westacott R E, Creek J L and Koh C A 2003 *J. Chem. Eng. Data* **48** 778
- [94] Botti A, Bruni F, Imberti S, Ricci M A and Soper A K 2003 *J. Chem. Phys.* **119** 5001
- [95] Bowron D T 2004 *Phil. Trans. R. Soc. B* **359** 1448
- [96] Botti A, Bruni F, Imberti S, Ricci M A and Soper A K 2004 *J. Chem. Phys.* **121** 7840
- [97] Botti A, Bruni F, Imberti S, Ricci M A and Soper A K 2004 *J. Chem. Phys.* **120** 10154
- [98] Botti A, Bruni F, Imberti S, Ricci M A and Soper A K 2005 *J. Mol. Liq.* **117** 77
- [99] Botti A, Bruni F, Imberti S, Ricci M A and Soper A K 2005 *J. Mol. Liq.* **117** 81
- [100] Ashbaugh H S, Patel H A, Kumar S K and Garde S 2005 *J. Chem. Phys.* **122** 104908
- [101] Buchanan P, Soper A K, Thompson H, Westacott R E, Creek J L, Hobson G and Koh C A 2005 *J. Chem. Phys.* **123** 164507
- [102] Imberti S, Botti A, Bruni F, Cappa G, Ricci M A and Soper A K 2005 *J. Chem. Phys.* **122** 194509
- [103] Milano G, Goudeau S and Müller-Plathe F 2005 *J. Polym. Sci. B* **43** 871
- [104] Pangburn T O and Bevan M A 2005 *J. Chem. Phys.* **123** 174904
- [105] Wu H J, Pangburn T O, Beckham R E and Bevan M A 2005 *Langmuir* **21** 9879
- [106] Soper A K 2005 *J. Phys.: Condens. Matter* **17** 3273
- [107] Pangburn T O and Bevan M A 2006 *J. Chem. Phys.* **124** 054712
- [108] Thompson H, Soper A K, Buchanan P, Aldiwan N, Creek J L and Koh C A 2006 *J. Chem. Phys.* **124** 164508
- [109] Belashchenko D K, Ginzburg A S and Mendelev M I 2000 *Russ. J. Phys. Chem.* **74** 577
- [110] Akkermans R L C and Briels W J 2001 *J. Chem. Phys.* **114** 1020
- [111] Belashchenko D K 2001 *Russ. J. Phys. Chem.* **75** 81
- [112] Belashchenko D K 2001 *Russ. J. Phys. Chem.* **75** 387
- [113] Belashchenko D K and Ginzburg A S 2001 *Russ. J. Phys. Chem.* **75** 790
- [114] Fukunaga H, Takimoto J and Doi M 2002 *J. Chem. Phys.* **116** 8183
- [115] Belashchenko D K, Polyanski R A and Pavlov R N 2002 *Russ. J. Phys. Chem.* **76** 454
- [116] Belashchenko D K 2002 *High Temp.* **40** 212
- [117] Belashchenko D K and Ginzburg A S 2002 *High Temp.* **40** 120
- [118] Belashchenko D K and Ostrovski O I 2006 *Inorg. Mater.* **42** 316
- [119] Bellisent-Funel M C, Chieux P, Levesque D and Weis J J 1989 *Phys. Rev. A* **39** 6310
- [120] Dzugutov M, Alvarez M and Lomba E 1994 *J. Phys.: Condens. Matter* **6** 4419
- [121] Rosenfeld Y and Kahl G 1997 *J. Phys.: Condens. Matter* **9** 89
- [122] Belashchenko D K 1998 *Russ. Metall.* **4** 126
- [123] Lobaskin V, Lyubartsev A P and Linse P 2001 *Phys. Rev. E* **63** 020401
- [124] Lyubartsev A P and Marcelja S 2002 *Phys. Rev. E* **65** 041202
- [125] Brunner M, Bechinger C, Strepp W, Lobaskin V and von Grünberg H H 2002 *Europhys. Lett.* **58** 926
- [126] Klein R, von Grünberg H H, Bechinger C, Brunner M and Lobaskin V 2002 *J. Phys.: Condens. Matter* **14** 7631
- [127] Lobaskin V, Brunner M, Bechinger C and von Grünberg H H 2003 *J. Phys.: Condens. Matter* **15** 6693
- [128] Reith D, Putz M and Müller-Plathe F 2003 *J. Comput. Chem.* **24** 1624
- [129] Lyubartsev A P, Karttunen M, Vattulainen P and Laaksonen A 2003 *Soft Mater.* **1** 121
- [130] Murtola T, Falck E, Patra M, Karttunen M and Vattulainen I 2004 *J. Chem. Phys.* **121** 9156
- [131] Lyubartsev A P 2005 *Eur. Biophys. J. Biophys. Lett.* **35** 53
- [132] Akkermans R L C 2000 *PhD Thesis* University Twente
- [133] Pangburn T O and Bevan M A 2005 *J. Chem. Phys.* **123** 174904
- [134] Bathe M and Rutledge G C 2003 *J. Comput. Chem.* **24** 876
- [135] Boutard Y, Ungerer P, Teuler J M, Ahunbay M G, Sabater S F, Perez-Pellitero J, Mackie A D and Bourasseau E 2005 *Fluid Phase Equilib.* **236** 25