

Data versus constraints in reverse Monte Carlo modelling: a case study on molecular liquid  
 $\text{CCl}_4$

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

2005 J. Phys.: Condens. Matter 17 S37

(<http://iopscience.iop.org/0953-8984/17/5/005>)

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 27/05/2010 at 20:18

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

# Data versus constraints in reverse Monte Carlo modelling: a case study on molecular liquid CCl<sub>4</sub>

Guillaume Evrard and László Pusztai

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

E-mail: lp@szfki.hu

Received 10 December 2004

Published 21 January 2005

Online at [stacks.iop.org/JPhysCM/17/S37](http://stacks.iop.org/JPhysCM/17/S37)

## Abstract

In reverse Monte Carlo modelling, experimental information (i.e. diffraction data) and *a priori* information (i.e. constraints introduced in the algorithm) are partly redundant. The extent of this redundancy for ('fixed neighbour') constraints determining the molecular geometry is studied systematically in the typical case of liquid CCl<sub>4</sub>. Results indicate that data with a very limited momentum transfer range are sufficient for deriving the intermolecular structure of such disordered systems when (intra)molecular geometry and intermolecular distances of closest approach are introduced by the appropriate algorithmic constraints.

## 1. Introduction

Reverse Monte Carlo (RMC) modelling is a computing method for deriving the structure of amorphous/disordered materials from diffraction data [1]. It has been successfully applied to a wide range of materials, from molecular liquids [2–4] to molten salts [5] and glasses [6]. (For a detailed description and more references, see the reviews that appear in this Special Issue [7, 8].)

The input (in part) and the output of an RMC simulation consist of a set of atoms and their positions, virtually put into a 'box' at the density corresponding to that of the material under study. The set of all conceivable such spatial configurations defines the parameter space upon which the RMC method operates. For a given configuration, the set of interatomic distances can be computed and binned into histograms in order to derive estimates of the partial pair correlation functions (PPCFs). These estimated PPCFs are sine-Fourier transformed to yield the partial structure factors (PSFs). The PSFs are then linearly combined, weighted according to the different scattering factors and concentration of every atomic species involved, to finally

give the ‘calculated’ diffracted intensity  $S_{\text{calc}}(Q)$ . The discrepancies between calculated and experimental data are measured by the usual  $\chi^2$  quantity:

$$\chi^2 = \sum_i \frac{[S_{\text{expt}}(Q_i) - S_{\text{calc}}(Q_i)]^2}{\sigma^2}, \quad (1)$$

where the sum is performed over all available data points  $\{Q_i, S(Q_i)\}$ , and possibly for several data sets.

The RMC technique is based on the Metropolis algorithm [9]: it is a random walk in the parameter space driven by the agreement between calculated and experimental data. Each step of the walk consists of a move of *one* atom of the configuration. Such a move modifies the distances histograms and thereby induces a change in the  $\chi^2$ . The move is accepted if the  $\chi^2$  decreases. The key of the Metropolis algorithm is, however, that if, on the contrary, the agreement between calculated and experimental data is decreased (i.e. the  $\chi^2$  increases), the move is still acceptable with the probability  $\exp[(\chi_{\text{old}}^2 - \chi_{\text{new}}^2)/2]$ . This random acceptance ensures that the algorithm does not get trapped in a local ‘ $\chi^2$  well’. That is, it prevents the algorithm from providing one ‘unique’ solution. The RMC method, indeed, yields a large number of possible solutions whose  $\chi^2$  value fluctuates around a common equilibrium value.

This lack of uniqueness (which has been the source of criticism of the RMC method) does not discredit the whole method, but it clearly indicates how to *use* RMC results. Conclusions from RMC simulations should be based on *collective* common features of the solutions, not on statistically insignificant, let alone unique, features of a resulting configuration.

The convergence of the Metropolis algorithm is based on statistical mechanics arguments, and is controlled by Boltzmann-like factors; therefore it yields a configuration with ‘maximum disorder’, which can correspond to unphysical solutions of the problem. In most cases, it is known that to some extent there is some order in the structure. In order to take into account this *a priori* information, constraints are added in the algorithm. Distances of closest approaches between atoms of a given type are a typical example of *a priori* physical knowledge introduced in the algorithm.

Diffraction data are not able to reveal whether there are molecules (i.e. fixed collections of atoms) in the sample under study: therefore, the molecular structure is usually a piece of information that needs to be added to RMC calculations as *a priori* physical knowledge. At present, most frequently flexible molecules are used (for a discussion on flexible versus rigid molecules, see [2]). Defining the geometry of flexible molecules is possible by means of the so-called ‘fixed neighbour constraints’ (FNCs [10]). FNCs work by fixing the neighbouring atoms (via their serial numbers in the configuration) around each particle. The lengths of the bonds defined in this way are allowed to vary within some tolerance. The extent of this tolerance will be of great importance throughout this study. Additional constraints such as coordination constraints and interatomic potentials can also be introduced [11] by adding some components to the  $\chi^2$  term that governs the acceptance/rejection of the random move.

These different *a priori* information sources, i.e. data and constraints, can be expected to be, to some extent, redundant. The present work is an attempt to identify and study some of these redundancies. It is practically impossible to investigate such behaviour of the RMC algorithm without being specific about given materials: the following results are based on simulations of liquid  $\text{CCl}_4$ , which is one of the most studied molecular liquids, using neutron diffraction data.

Previous work [12] indicates that the medium and long ranges of  $Q$  are redundant for cases of some simple liquids and covalent and metallic glasses, and that distances of closest approach have no significant effects. In the present study, we focus on the possible redundant information content of the FNCs and of the high  $Q$  end of the diffraction data range of  $Q$ .

**Table 1.** The data range and FNCs for the first nine runs with CCl<sub>4</sub>.

FNC (Å) ⇒	C–Cl ∈ [1.71, 1.85]	C–Cl ∈ [1.69, 2.00]	C–Cl ∈ [1.69, 2.19]
<i>Q</i> range (Å <sup>-1</sup> ) ↓	Cl–Cl ∈ [2.7, 3.1]	Cl–Cl ∈ [2.7, 3.3]	Cl–Cl ∈ [2.7, 3.5]
[0.55, 9.15]	A1	B1	C1
[0.55, 6.9]	A2	B2	C2
[0.55, 4.975]	A3	B3	C3

**Table 2.** Common run parameters for the first nine runs with CCl<sub>4</sub>.

Number of atoms	10 240 (2048 molecules)
Density	0.0319 Å <sup>-3</sup>
Cubic cell size, half-edge	34.235 Å
<i>r</i> spacing in distance histograms	0.1 Å
Move amplitude for all atoms	0.1 Å
Distances of closest approach:	
C–C	3.3 Å
C–Cl	1.69 Å
Cl–Cl	2.7 Å
$\sigma$ (data standard deviation)	0.001

It should be remembered that FNCs define the (intra)molecular structure; therefore, it is the *intermolecular* correlations that will be investigated as a function of the FNCs and the  $Q_{\max}$  value of the data modelled.

## 2. RMC simulation runs with CCl<sub>4</sub>

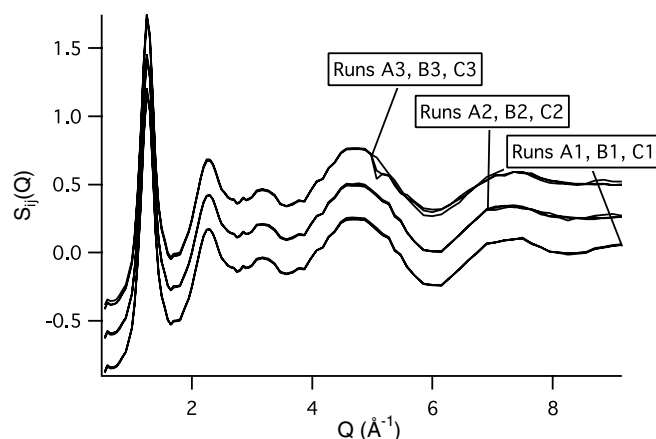
Liquid CCl<sub>4</sub> provides a good example. It is a simple system with weak intermolecular interactions, for which relevant valuable information can be retrieved from diffraction measurements (see e.g. [2]). Besides this, its tetrahedral shape makes it particularly well suited for description with FNCs. The neutron diffraction data used here were originally taken at the Budapest Research Reactor for a previous study [3].

The FNCs define the molecular geometry (via the *intramolecular* distances between specific atoms of the configuration); one can thus expect them to also strongly constrain the high  $Q$  part of the diffraction data (which part, for disordered materials, is dominated by short range correlations).

To check possible redundancies or competing effects between the high  $Q$  range and FNCs, several RMC runs were performed with constraints given by a 3 FNC  $\times$  3  $Q$  range constraints grid (table 1) defined as follows: the original neutron diffraction data were measured up to 9.15 Å<sup>-1</sup> (tabulated in 100 points), from which two restricted data sets were subsequently built by chopping off the high  $Q$  end at 6.9 Å<sup>-1</sup> (85 points) and 4.975 Å<sup>-1</sup> (70 points). Similarly three different sets of FNCs were defined from ‘strict’ to ‘loose’, allowing the shape of the CCl<sub>4</sub> molecule to depart significantly from the ideal tetrahedron.

Note that since atoms are moved *one by one* in the basic step of the RMC random walk, the algorithm requires the molecule to be flexible to a certain extent. There is an obvious trade-off between the move amplitude and the allowed deviation from some ideal molecular geometry. It is, however, possible to use the RMC scheme for *rigid* molecules by defining the random step as a custom displacement of the *whole* molecule [7].

The otherwise common parameters for the nine RMC runs appear in table 2.



**Figure 1.** Modelled versus total structure factors. Bottom: runs A1, B1 and C1 ( $Q$  range up to  $9.15 \text{ \AA}^{-1}$ ). Middle (vertical shift of 0.25): runs A2, B2 and C2 ( $Q$  cut-off at  $6.9 \text{ \AA}^{-1}$ ). Top (vertical shift of 0.5): runs A3, B3 and C3 ( $Q$  cut-off at  $4.975 \text{ \AA}^{-1}$ ). The TSFs can be recovered beyond the cut-off, although some slight but sharp departure from the experimental curve can be seen (at the detailed level) at the cut-off.

It must be noted that in the RMC method, the  $\sigma$  value, which enters the definition of the  $\chi^2$  and which is in principle defined by experimental uncertainties, is in fact one adjustable parameter that controls the fit quality and possibly weights the different data sets. For example, a previous attempt with a larger value of  $\sigma$  (0.005) failed to produce good fits to the data.

Each calculation was started from an identical initial disordered configuration. The duration of one run was typically 20 or 24 h (for one processor, on a two-processor 700 MHz PC running under Linux, with the RMC ++ implementation of the RMC method [7]). The number of accepted moves per atom varied between 170 and 340 and, in all cases, convergence was ensured by the lack of further evolution of the  $\chi^2$  parameter.

The different RMC outputs can be visualized by examining the total structure factors (TSF) modelled, the resulting partial structure factors, as well as the PPCFs. We must stress however that the configuration contains much more information than just the PPCFs, and that additional tools (such as bond angle distributions) are easily accessible.

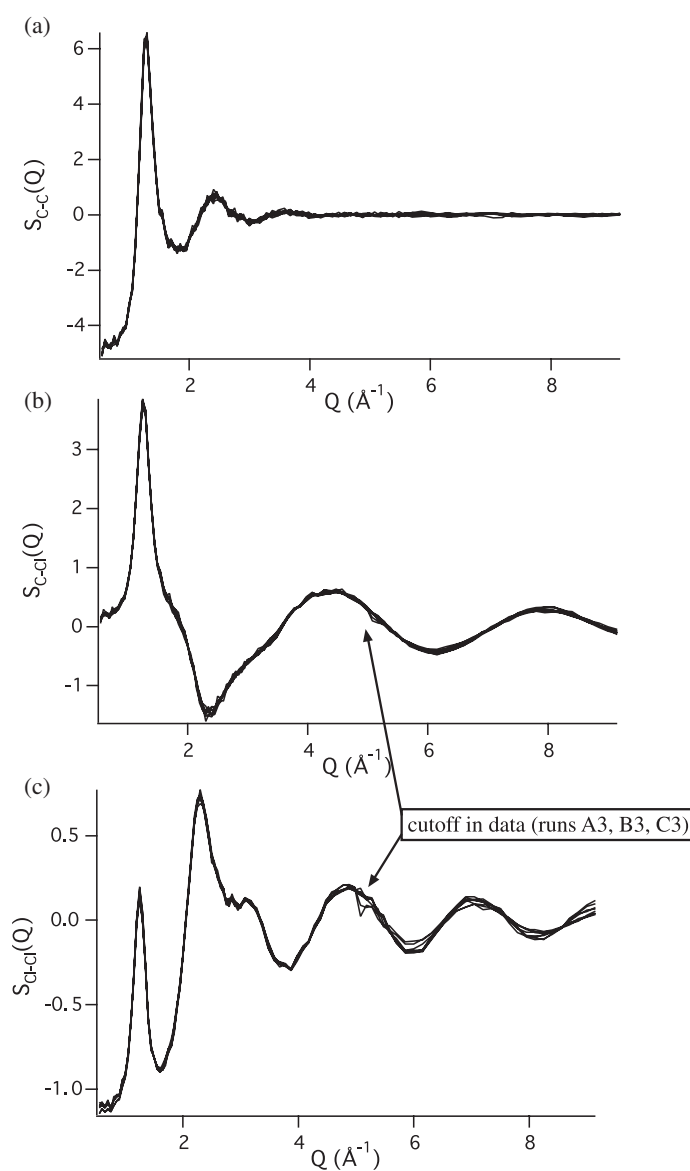
Before going into detailed discussions, it should be mentioned that in most of the cases studied here, the FNCs will dominate sharp features of the PPCFs (which appear in the  $r$  ranges corresponding to *intramolecular* distances) since the widths of the FNCs are smaller than, or comparable to, the peak broadening due to  $Q_{\text{max}}$ : the data will not be able to make sharp features sharper than the FNCs. However, since it is the *intermolecular* correlations that are of interest in a ( $n$  RMC method aided) diffraction study, the above limitation is not particularly important from the point of view of the current study.

### 2.1. Information in $Q$ space

The goodness of fit of the modelling appears in figure 1.

Since the agreement between calculated and experimental data is a requirement of the RMC method, the fits are rather good. Noticeably, the features of the TSF are recovered beyond the cut-offs, although detailed examination reveals a slight but sharp departure of the ‘calculated’ data from the experimental curve where the  $Q$  range has been shortened.

Figure 1 indicates clearly that the information content of the high  $Q$  part of the structure factor is somehow redundant, with the information content of its lower  $Q$  part



**Figure 2.** Partial structure factors from the nine RMC runs. (a): C–C; (b): C–Cl; (c): Cl–Cl. All sets are quite homogeneous; only the Cl–Cl PSF exhibits noticeable changes due to the different constraints applied.

combined with the intrinsic *a priori* information of the model (i.e. FNCs and very basic modelling assumptions). Otherwise, the departures of the calculated structure factors from the experimental data would certainly be much larger in the chopped-off region.

One can ask whether the slight changes observed when experimental and/or algorithmic constraints are loosened for the *total* structure factor are equally shared by the *partial* structure factors or whether, on the contrary, the loosening affects the partial structure factors in different ways. Figure 2 shows the three partials for all sets. As can be expected from the small variations

between the different *total* structure factors, *partial* structure factors (PSFs) do not exhibit large changes, either. Interestingly, the different cut-offs in the  $Q$  range visible in the TSF seem to originate only in the Cl–Cl partial. In other words, the suppressed information contained in the high  $Q$  range mainly concerns the Cl–Cl distance. This is in agreement with the fact that, in general, the high  $Q$  part of the structure factor describes intramolecular features, of which the Cl–Cl distance is the most flexible. (Note also that the Cl–Cl partial has the largest weight in the total neutron structure factor and, therefore, any effect can be best seen for this function. However, this cannot be taken as a full explanation since the C–Cl partial, which also has a significant weight, is totally unaffected.)

## 2.2. Information in $r$ space

Since the pair correlation functions  $g(r)$  (and the different partials) are related to the structure factor by a Fourier-transformation-like process, it is possible that the loss of information due to the suppression of the high end of the  $Q$  range in the structure factor, which is only marginally visible in  $Q$  space, might appear in  $r$  space.

It must be stressed that partial functions  $g(r)$  are always calculated from the configuration obtained by the RMC method and NOT by inverse Fourier transformation of the  $S(Q)$ . In this way, we remain free from the usual problems occurring when applying Fourier transforms to data limited in  $Q$  range that plague direct methods of inversion.

Figure 3 shows the three partials, which, by and large, are almost identical.

No feature distinguishes the different C–C PPCFs even on a detailed scale. This means that all the information needed to recover this PPCF is contained in the loosest constraints combined with experimental data over the shortest  $Q$  range. In other words, the distribution of the molecular centres is well defined by the data, and is not affected by the rigidity of the molecular geometry.

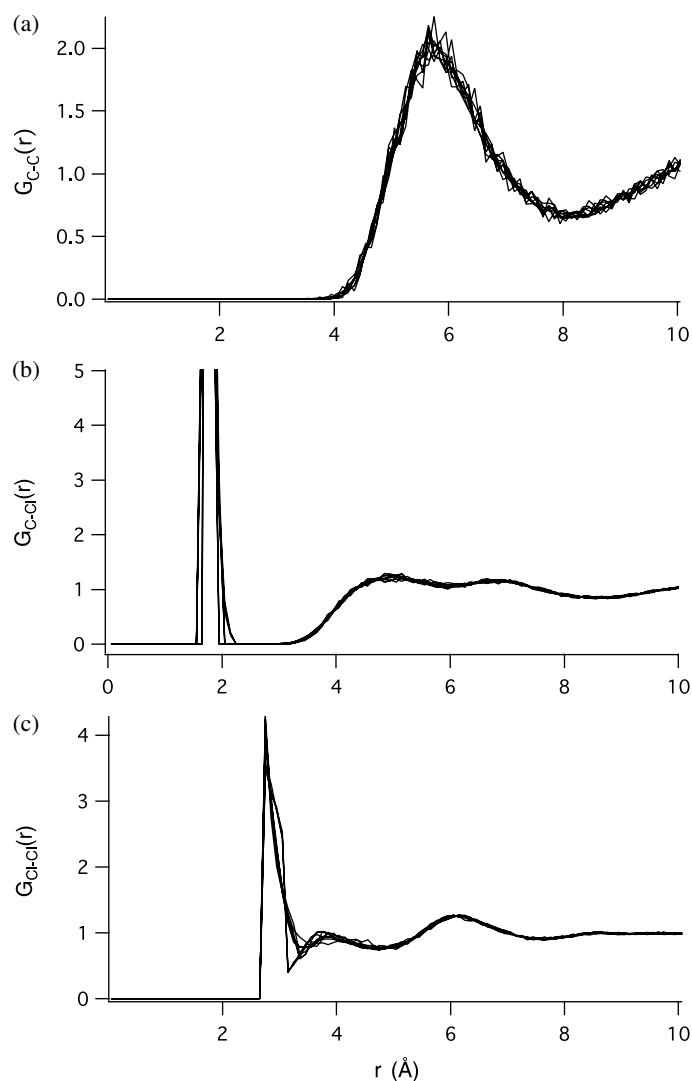
The C–Cl partials can be distinguished only by the details of the intramolecular peak which, as a result of loosening FNCs, is broadened slightly. The  $r$  spacing was set to 0.1 Å and at this level no additional information can be expected from the data for the  $Q$  ranges considered. Therefore, the widening of the intramolecular features is just the effect of the release of the molecular geometrical constraints, the RMC method yielding the ‘most disordered’ configuration given the constraints.

The Cl–Cl partials can also be separated at the detailed level only (figure 4). When the molecular geometry is rigid, the intramolecular Cl–Cl peak is clearly separated from the first intermolecular peak (at  $\sim 3.8$  Å). Loosening the FNCs allows the two contributions to become mixed, and it becomes impossible to disentangle them clearly (although the small maximum at 3.8 Å never disappears).

In general, the  $g(r)$  has the sharpest features for the full extent of the data and the tightest constraints. Suppressing the high  $Q$  bits of the  $S(Q)$  has a minor effect, unless it is accompanied by a release of the FNCs, which appear to be the driving features.

## 2.3. Additional RMC runs

In a second series of calculations, we tried to identify the structural information specifically contained in the upper part of the  $Q$  region. For this purpose, we produced a set of artificial data extending to a (very) large maximum  $Q$  value (up to  $40 \text{ \AA}^{-1}$ ), from the resulting configuration of run A1 (i.e. obtained with the strictest constraints and the full set of experimental data). A series of RMC runs was subsequently performed using the loosest FNCs defined previously, and again suppressing progressively the high end of the  $S(Q)$  range. Two additional simulations



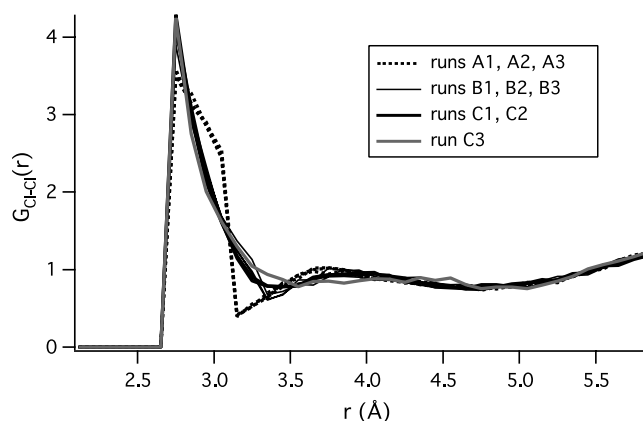
**Figure 3.** Partial pair correlation functions, from the nine sets of RMC runs. (a): C–C; (b): C–Cl; (c) Cl–Cl. On a large scale they are all equivalent. Detailed examination is required in order to separate the different curves.

without data, but with the loosest and the tightest set of FNCs used previously, were also made. Table 3 summarizes the different input of this second series.

On the whole, this second set of simulations confirms results described in the previous section: the information gained by extending the  $Q$  range in the  $S(Q)$  data is only marginal. This is particularly true in the region of large values of the maximum  $Q$  (say, beyond about  $15 \text{ \AA}^{-1}$ ).

The different C–C PPCFs cannot be distinguished (and, therefore, they are not shown). The different Cl–Cl PPCFs are also very close to one another; only the one obtained with the cut-off at  $10 \text{ \AA}^{-1}$  can be separated from the others (see figure 5(b)). A similar conclusion can be drawn concerning the C–Cl PPCF (figure 5(a)) but it is even less visible.





**Figure 4.** A detailed view of the Cl–Cl partial pair correlation function. At this particular point (just after the intramolecular peak), tightening the FNCs allows one to disentangle the intramolecular and the intermolecular contributions to the 3.8 Å ‘bump’.

**Table 3.** Input for the second series of RMC runs with  $\text{CCl}_4$ : ‘loose’ FNCs for runs with data and ‘strict’ and ‘loose’ FNCs for ‘hard sphere’ runs (without data).

Run	$Q$ range ( $\text{\AA}^{-1}$ )	C–Cl FNC ( $\text{\AA}^{-1}$ )	Cl–Cl FNC ( $\text{\AA}^{-1}$ )	$\Delta r$ step ( $\text{\AA}$ )	Number of points
D1	[0.55, 40]	[1.69, 2.19]	[2.7, 3.5]	0.031	409
D2	[0.55, 30]	[1.69, 2.19]	[2.7, 3.5]	0.041	309
D3	[0.55, 25]	[1.69, 2.19]	[2.7, 3.5]	0.05	259
D4	[0.55, 20]	[1.69, 2.19]	[2.7, 3.5]	0.062	209
D5	[0.55, 15]	[1.69, 2.19]	[2.7, 3.5]	0.083	159
D6	[0.55, 10]	[1.69, 2.19]	[2.7, 3.5]	0.1	109
HS0	No data	[1.71, 1.85] Å	[2.7, 3.1]	0.1	0
HS1	No data	[1.69, 2.19] Å	[2.7, 3.5]	0.1	0

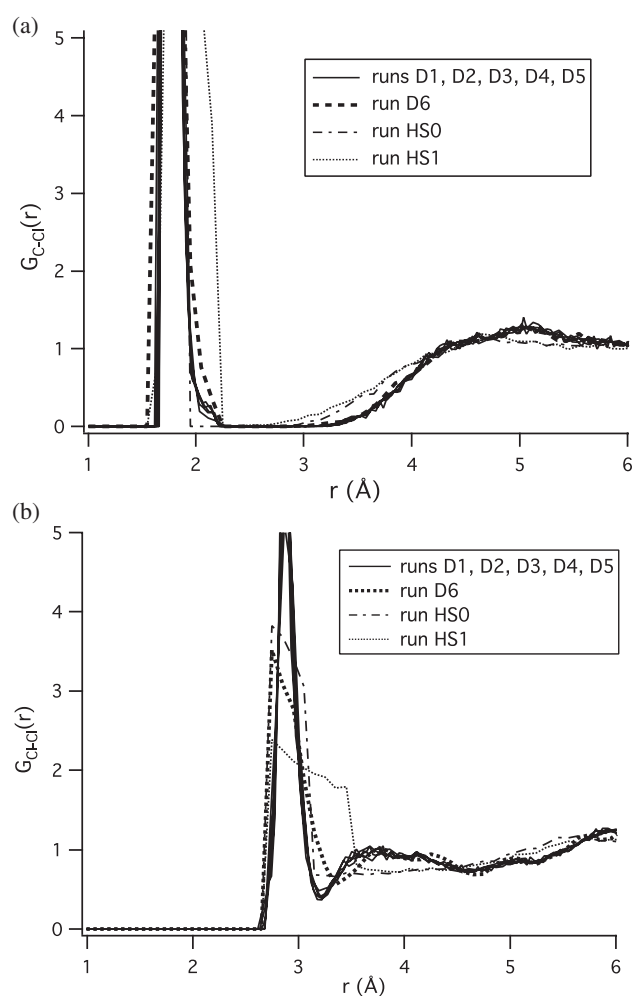
The additional information contained in the higher  $Q$  extensions of the structure factor is manifested in the sharpness of the functions  $g(r)$ . Since the sharpest features are the intramolecular peaks, that is where all the supplementary data are transposed in  $r$  space. No other significant change in the PPCFs was found; that is, intermolecular characteristics (corresponding to orientational correlations) were not affected at all.

Finally, since most of the relevant information contained in the  $S(Q)$ , not redundant with the RMC assumptions and constraints, seems to be encompassed in the shortest  $Q$  range considered in this study (i.e. below  $4.7 \text{ \AA}^{-1}$ ), we made attempts to run RMC simulations *without data* (applying the FNC constraints only). Results indicate that the density constraints define the C–C distance (not shown). For the C–Cl partial, the main intermolecular features are roughly reproduced with constraints only. In contrast, for the Cl–Cl partial the data are needed to obtain the main features of the PPCF at medium range.

### 3. Summary and conclusion

The results described above confirm earlier findings [12] (obtained with distances of closest approach only) and indicate that diffraction data of limited momentum transfer range can be used to derive the structure of disordered materials with RMC modelling.

The high  $Q$  parts of the structure factor correspond to sharp features in  $r$  space. For disordered materials, these features appear at low  $r$  values only (at medium  $r$  range and beyond,



**Figure 5.** Detailed views of the C–Cl (a) and Cl–Cl (b) PPCFs. The decrease of the sharpness of the intramolecular peak is the only visible feature that derives from the variation of the  $Q$  range of the structure factor in the input of the RMC simulation.

smoothing due to disorder becomes prominent). For molecular systems, this low  $r$  region contains—mostly—intramolecular distances, which, in a reverse Monte Carlo calculation, are determined by the FNCs. Consequently, high  $Q$  information is not essential for deriving structural information at ‘intermediate’ (*intermolecular*) ranges. We deliberately avoid giving numbers for distances corresponding to ‘short’, ‘intermediate’ and ‘long’ range, since they would depend on the materials considered.

Furthermore, at ‘short’  $r$  range (corresponding roughly to molecular size), RMC constraints such as distances of closest approach and FNCs provide essential pieces of information competing with the high  $Q$  part of the diffraction data (or replacing it, if it is not available). However, due to the structure of the algorithm, constraints are given an overwhelming weight. High momentum transfer may nevertheless indicate very fine structure, if the  $Q$  range is wide enough to correspond to  $r$  resolution smaller than the distance ranges defined by FNCs (e.g. the detailed shape of the intramolecular peak). The relevance of such details, however, remains to be assessed, as are the uncertainties in the RMC modelling output.

In the interesting cases where the molecular structure makes an important contribution to the definition of intermediate range order, RMC constraints and diffraction data are competing/redundant. On one hand, this can be used to disentangle intermolecular from intramolecular features in the functions  $g(r)$ ; it can also help to detect systematic errors in diffraction data, or inconsistencies between different data sets. But, on the other hand, it must be kept in mind that the implementation of some RMC constraints (such as FNCs, coordination constraints and potentials) is partly arbitrary. Consequently, all RMC results must always be given with the set of applied constraints.

Finally, on the basis of the findings of the current study, it must be stressed again that the RMC method must never be used as a *black box* and that, eventually, its results must always be scrutinized and validated or discarded on the basis of physics or chemistry arguments.

### Acknowledgments

This work was supported by the EU ‘Centre of Excellence’ grant EU-ICAI-CT-2000-70029. LP was also funded by the ‘Hungarian Basic Research Fund’ (OTKA), through grants Nos T32308 and T42495.

### References

- [1] McGreevy R L and Pusztai L 1988 *Mol. Simul.* **1** 359
- [2] Pusztai L and McGreevy R L 1997 *Mol. Phys.* **90** 533
- [3] Jóvári P, Mészáros G, Pusztai L and Sváb E 2001 *J. Chem. Phys.* **114** 8082
- [4] Jóvári P 1999 *Mol. Phys.* **97** 1149
- [5] McGreevy R L and Pusztai L 1990 *Proc. R. Soc. A* **430** 241
- [6] Keen D A and McGreevy R L 1990 *Nature* **334** 423–5
- [7] See Evrard G and Pusztai L 2005 *J. Phys.: Condens. Matter* **17** S1 (software and documentation available from [www.szfki.hu/~nphys/](http://www.szfki.hu/~nphys/))
- [8] Keen D A, Tucker M G and Dove M T 2005 *J. Phys.: Condens. Matter* **17**
- [9] Metropolis N A, Rosenbluth A W, Teller E and Teller J 1953 *J. Chem. Phys.* **21** S15
- [10] Pusztai L and McGreevy R L 1997 *NFL Studsvik Annual Report for 1996* OTH:21
- [11] McGreevy R L 1997 *RMCA Manual* available on [www.studsvik.uu.se](http://www.studsvik.uu.se)
- [12] Gereben O and Pusztai L 1995 *Phys. Rev. B* **51** 5768