RMCA

Version 3.14

June 2005

1. Introduction

One of the major problems in the analysis of diffraction data on disordered systems has been the lack of any general method for producing structural models that agree quantitatively with the data. Most analysis is extremely qualitative and based on a few features of the data, for example peak positions and coordination numbers derived from radial distribution functions. Monte Carlo and Molecular Dynamics simulations based on an interatomic potential sometimes agree well with experiment (though comparison is normally made with radial distribution functions rather than structure factors), but usually the agreement is only qualitative and occasionally there are major differences. It is not obvious in most cases how the potential should be altered to improve the level of agreement; an iterative procedure is computationally extremely expensive and has only been applied in one or two instances.

The reverse Monte Carlo (RMC) method [4] overcomes these problems. It is a method for producing three dimensional models of the structure of disordered materials that agree quantitatively with the available diffraction data. No interatomic potential is required and data from different sources (neutrons, X-rays, EXAFS) may be combined. The structural model is actually fitted to the data and so there must be good agreement (given that the data do not contain significant systematic errors).

RMCA is a general purpose Fortran code for reverse Monte Carlo (RMC) modelling. This manual first describes the RMC method and then goes on to describe the RMCA program.

1.1 Changes from version 3.12

- Some bugs causing problems with modelling G(r) and coordination constraints were fixed
- Average and r.m.s bond valence sum constraints added, see 5 and 10.3.12.
- Minimum radius parameter for triplet constraints added, see 10.3.15.

2. Metropolis Monte Carlo (MMC)

RMC is a variation of the standard MMC procedure [1]. For those unfamiliar with such Monte Carlo methods it is useful to first introduce MMC. The principle is that we wish to produce a statistical ensemble of atoms (configuration) with a Boltzmann distribution of energies. Rather than simply generating and sampling configurations completely at random, which would be a very inefficient procedure, we make use of a weighted sampling procedure (Markov chain) that satisfies certain requirements.

(a) variables $x^{(n)}$ are generated following a rule that requires the $(n + 1)^{nt}$ element to have a probability distribution $x^{(n+1)}$ that is only dependent on the distribution $x^{(n)}$ of the nth element.
(b) If $P(x => y)$ is the probability of reaching state y from state x, then $P(x => y)$ must permit movement to every state in the ensemble.
(c) Microreversibility must be satisfied, that is $xP(x => y) = yP(y => x)$ when the system is in equilibrium.

For an ensemble in which the number of particles, volume and temperature are fixed (NVT) this may be achieved by the following algorithm.

1. N atoms are placed in a cell with periodic boundary conditions, by which we mean that the cell is surrounded by images of itself. Normally a cubic cell is used, although other geometries may be chosen. For a cube of side L the atom number density $\rho = N/L3$ must equal the required density of the system. The probability of this particular configuration (old = o) is given by

$$P_o \propto \exp\left(- U_o / kT \right) \tag{1}$$

where U_o is the total potential energy, which may be calculated on the basis of a specified form of the interatomic potential, and T is the specified temperature.

2. One atom is moved at random. The probability of the new (n) configuration is

$$P_n \propto \exp\left(- U_n / kT\right) \tag{2}$$

and hence

$$P_n / P_o = \exp\left(-(U_n - U_o)/kT\right) = \exp\left(-\Delta U / kT\right) \tag{3}$$

3. If $\Delta U < 0$ the new configuration is accepted and becomes the next starting point.
If $\Delta U > 0$ it is accepted with probability P_n/P_o otherwise it is rejected and we return to the previous configuration.
4. The procedure is repeated from step 2.

As atoms are moved U will decrease until it reaches an equilibrium value about which it will then oscillate. The maximum size of the random move is normally adjusted so that the ratio of accepted to rejected moves in equilibrium is approximately unity. Configurations are considered to be statistically independent when separated by at least N accepted moves and are then saved. In this way an appropriate ensemble is generated.

3. RMC - the basic method

In RMC we assume that an experimentally measured structure factor $A^E(Q_i,)$ contains only statistical errors that have a normal distribution (this will be discussed in more detail later). The difference between the real structure factor, $A^C(Q_i)$, which can be calculated from a model of the real structure, and that measured experimentally is then

$$e_i = A^C(Q_i) - A^E(Q_i) \tag{4}$$

and has probability

$$p(e_i) = \frac{1}{\sqrt{2\pi}\sigma(Q_i)} \exp\left(-\frac{e_i^2}{2\sigma(Q_i)^2}\right) \tag{5}$$

where $\sigma(Q_i)$ is the standard deviation of the normal distribution.

The total probability of A^C is

$$P = \prod_{i=1}^{m} p(e_i) = \left(\frac{1}{\sqrt{2\pi}\,\overline{\sigma}}\right)^m \exp\left(-\sum_{i=1}^{m} \frac{e_i^2}{2\sigma(Q_i)^2}\right) \tag{6}$$

where m is the number of Q_j points in A^E and

$$\overline{\sigma} = \left(\prod_{i=1}^{m} \sigma(Q_i)\right)^{1/m} \tag{7}$$

In order to model the structure of a system using A^E we therefore wish to create a statistical ensemble of atoms whose structure factor satisfies the above probability distribution. Writing the exponent as

$$\chi^2 = \sum_{i=1}^{m} \left(A^C(Q_i) - A^E(Q_i)\right)^2 / \sigma(Q_i)^2 \tag{8}$$

then $P \propto \exp(-\chi^2/2)$ and it can immediately be seen that $\chi^2/2$ in RMC is equivalent to U/kT in MMC. The algorithm for RMC is therefore as follows.

4

1. Start with an initial configuration with periodic boundary conditions. The positions of the N atoms may be chosen randomly, they may have a known crystal structure, or they may be a configuration from a different simulation or model.

2. Calculate the radial distribution function for this old configuration

$$g_o^C(r) = \frac{n_o^C(r)}{4\pi r^2 \Delta r \rho} \tag{9}$$

where $n_o^C(r)$ is the number of atoms at a distance between r and $r + \Delta r$ from a central atom, averaged over all atoms as centres and ρ is the number density. The configuration size L should in principle be sufficiently large that there are no correlations across the cell, so that $g(r > L/2) = 1$. The radial distribution function $g(r)$ is only calculated for $r < L/2$ and the nearest image convention is used to determine the atomic separations.

3. Transform to the total structure factor

$$A_o^2(Q) - 1 = \frac{4\pi\rho}{Q} \int_0^\infty r\left(g_0^C(r) - 1\right) \sin Qr \, dr \tag{10}$$

where Q is the momentum transfer.

4. Calculate the difference between the measured total structure factor $A^E(Q)$ and that determined from the configuration $A^C(Q)$

$$\chi_o^2 = \sum_i \left(A_o^C(Q_i) - A^E(Q_i)\right)^2 / \sigma(Q_i)^2 \tag{11}$$

where the sum is over the m experimental points and $\sigma(Q_i)$ is the experimental error. In practice a uniform σ is normally used, since the distribution of systematic errors is unknown.

5. Move one atom at random. Calculate the new radial distribution function, $g_n^C(r)$ and total structure factor $A_n^C(Q)$, and

$$\chi_n^2 = \sum \left(A_n^C(Q_i) - A^E(Q_i)\right)^2 / \sigma(Q_i)^2 \tag{12}$$

6. If $\chi_n^2 < \chi_o^2$ the move is accepted and the new configuration becomes the old configuration. If $\chi_n^2 > \chi_o^2$ then the move is accepted with probability $\exp(-(\chi_n^2-\chi_o^2)/2)$. Otherwise it is rejected.

7. Repeat from step 5.

As this process is iterated χ^2 will initially decrease until it reaches an equilibrium value about which it will fluctuate. The resulting configuration should be a three dimensional structure that is consistent with the experimental total structure factor within the experimental error. Statistically independent configurations may then be collected. In MMC configurations are normally assumed to be independent if separated by N accepted moves but in RMC we usually use at least $5N$.

The algorithm used here is not strictly statistically correct, since we are actually sampling χ^2 (by varying A^C) and not the data (by multiple measurements of A^E).
We should therefore use

$$\left(\frac{\chi_n^2}{\chi_o^2}\right)^{\frac{m}{2}-1} \exp\left(-\left(\chi_n^2 - \chi_o^2\right)/2\right) \tag{13}$$

in place of $\exp(-(\chi_n^2-\chi_o^2)/2)$. However, given that most experimental data do not contain only statistical errors, and the loss of direct analogy with MMC, the former algorithm has been preferred.

The distinction between RMC and MMC is simply that in RMC the difference between calculated and measured total structure factors (χ^2) is sampled, while in MMC the potential energy U is sampled. Otherwise the two algorithms are identical. It is particularly important that RMC uses a proper Markov chain, so that

the final structure should be independent of the initial configuration. This makes the method an *ab initio* structural determination, rather than a refinement. However in some circumstances the method is deliberately used as a refinement. This involves only accepting moves that decrease χ^2 (i.e. $\sigma \to 0$)and corresponds to setting $T = 0$ in MMC.

Similar algorithms have been tried by others. The earliest examples are the work of Averbach and colleagues [2,3] who used models of only a few hundred atoms (all that could then be managed on the computers available), modelled *g(r),* and used converging moves only (i.e. χ^2 must always decrease). This makes the results entirely dependent on the starting configuration and so the method did not find general applicability. In a later example Bertagnolli and colleagues [7] used an algorithm very similar to that for RMC, but only modelled the inter-molecular part of *g(r)* for a series of molecular liquids.

4. Multiple data sets

The algorithm described in the previous section is specifically for modelling a single set of diffraction data, which could be obtained, using either X-rays, neutrons, or electrons. The fit may be either to the structure factor or to the radial distribution function, though the former is to be preferred because the distribution of errors in the latter may be highly non-uniform. In practice a fit may be made first to the radial distribution function, then to a subset of the total structure factor points, before being made finally to all the structure factor points. This considerably reduces the time required.

The RMC method is more general than this simple algorithm in that any set or sets of data which can be directly calculated from the structure can be modelled. It can be applied to isotopic substitution in neutron diffraction, or equivalently to anomalous scattering in X-ray diffraction, to EXAFS and possibly to NMR data. All data sets can be modelled simultaneously by adding the respective χ^2 values.

For a multicomponent system where the fit is to several different total structure factors (indicated by index *k)* we have

$$\chi^2 = \sum_k \chi_k^2 = \sum_k \sum_{i=1}^{m} \left(F_k^C(Q_i) - F_k^E(Q_i) \right)^2 / \sigma_k(Q_i)^2 \tag{14}$$

For neutron diffraction

$$F_k^E(Q_i) = \sum_\alpha \sum_\beta c_\alpha c_\beta \bar{b}_{\alpha k} \bar{b}_{\beta k} \left(A_{\alpha\beta}(Q_i) - 1 \right) \tag{15}$$

where c_α is the concentration and $b_{\alpha k}$ the coherent scattering length for species α in sample *k*. $A_{\alpha\beta}(Q_i)$ are the partial structure factors. For X-ray diffraction

$$F_k^E(Q_i) = \sum_{\alpha\beta} c_\alpha c_\beta f_{\alpha k}^*(Q_i) f_{\beta k}^*(Q_i) \left(A_{\alpha\beta}(Q_i) - 1 \right) \tag{16}$$

where

$$f_{\alpha k}^*(Q_i) = \frac{f_{\alpha k}(Q_i)}{\sum_\alpha c_\alpha f_{\alpha k}(Q_i)^2} \tag{17}$$

$f_{\alpha k}(Q_i)$ is the Q dependent form factor for X-rays of wavelength λ_k. The normalised value $f^*_{\alpha k}(Q_i)$ is used in the definition of $F^E_k(Q_i)$ for the usual case of the scattered intensity being measured with constant statistical error. For EXAFS data

$$F_\alpha^E(Q_i) = \sum_\beta 4\pi \rho \int r^2 \left(g_{\alpha\beta}(r) - 1 \right) f_\beta(Q_i, r) dr \tag{18}$$

where we have replaced *k* with a in $F^E(Q_i)$ because the spectrum is measured at the absorption edge of species α. Here $f_\beta(Q_i,r)$ is the contribution to the EXAFS spectrum of a single atom of type β at a distance *r*

which is calculated by one of the standard EXAFS data analysis packages such as the SERC Daresbury program EXCURV92. When fitting EXAFS data it is possible to *Q*-weight the spectra.

For simultaneous fitting of data sets obtained by different experimental techniques the separate χ^2 values are simply summed to give one value. The relative weighting of the different data sets is determined by the choice of the various a values. Clearly the required computer time increases significantly if multiple data sets are fitted.

5. Constraints

Other information that cannot be used directly can be made use of in the form of constraints; this may include NMR, EPR, Raman scattering and chemical knowledge. The most commonly used constraint is on the closest distance of approach of two atoms. Because of systematic errors in the experimental data, and often because of the limited data range, the data would not forbid some atoms from coming very close together. However we know that this is physically unrealistic so an excluded volume is defined. Often realistic values for the closest approach distances can be determined from direct Fourier transformation of the measured total structure factors. It is usually obvious if an unsuitable choice has been made as spurious sharp spikes occur in g(*r*) at low *r*.

While the closest distance of approach constraint is very simple, it is very powerful when used in conjunction with a fixed density. For many materials the dominant effect determining the structure is packing, and hence to implicitly include information on atomic sizes in the model (these are minimum sizes rather than, for example, ionic radii) severely limits the number of structures that are consistent with the data.

The second most commonly used constraint is on the coordination of the atoms. A coordination number $n_{\alpha\beta}$ is defined as being the number of atoms of type β between two fixed distances of one of type α. Normally the lower fixed distance is the closest distance of approach of the two types of atom (or equivalently zero). If we define the proportion of atoms of type α in the configuration with a particular coordination as f_{RMC} and the desired proportion with such a coordination as f_{req} then we can add an additional term to χ^2:

$$\chi^2 = \ldots + \left(f_{req} - f_{RMC}\right)^2 / \sigma_c \tag{19}$$

Obviously multiple coordination constraints can be applied by adding additional terms. The parameter σ_c, in this case simply acts as a weighting of the coordination constraint relative to the data. If $\sigma_c \approx 0$ it is effectively impossible for atoms with the constrained coordination to change it; this can be used to mimic the effect of covalent bonding. In many cases hard sphere Monte Carlo simulation with such coordination constraints, that is RMC with no data, can be used to produce structures with suitable topology prior to fitting the data.

This is a constraint on coordination numbers of individual (although unspecified) atoms. It is also possible to constrain average coordination numbers, and their r.m.s. values, in the same way. Average coordination numbers can be obtained from EXAFS data, and using these as constraints rather than fitting directly to the spectra is an alternative way of using such data.

A third constraint type is the angular constraint for triplets of atoms (see figure below). An angle $\boldsymbol{\theta}$ is defined by the three atoms $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ which can be of the same type or different types. An extra term to χ^2 can be expressed as:

$$\chi^2 = \ldots + \sum_i \left(\theta_i - \theta_{pref}\right)^2 / \sigma_T \tag{20}$$
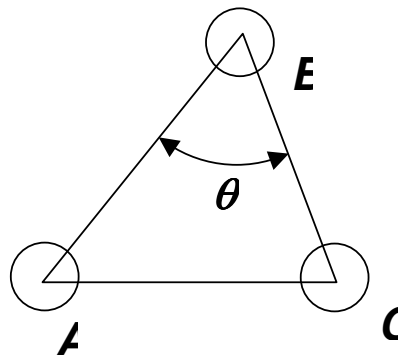
where θ_{pref} is the preferred angle and σ_T serves as a weighting of the angular constraint relative to data and coordination constraints. A small σ_T give a sharp angular distribution (small width) and a large σ_T gives a wide angular distribution around $\theta_{pref.}$

Bond valence sum (BVS) [48] calculations can be used in to constrain the average and root mean square BVS of one or more particle types. These constraints, based on a pseudo-potential approach, are very useful e.g. to obtain reasonable nearest neighbour environments in structures with high degree of disorder due to vacancies. For each BVS R.M.S. constraint a term,

$$\chi^2 = ... + \sum_j \left(v_j - \langle v \rangle\right)^2 / \sigma_{val}^2 \tag{21}$$

is added to the total χ^2. Here $< v >$ is the actual average BVS and the sum runs over all particles j of a type. In the case of average BVS constraints, a term

$$\chi^2 = ... + \left(aval_{req} - \langle v \rangle\right)^2 / \sigma_{aval}^2 \tag{22}$$

is added to the total χ^2, where $< v >$ and avalreq are the actual and desired average BVS's resp. In both cases the individual BVS for particle j is calculated either as

$$v_j = \sum_i \left(R_0 / R_i^j\right)^N \text{ or } v_j = \sum_i \exp\left(\frac{R_0 - R_i^j}{B}\right) \tag{23}$$

where the sum runs over all bonds with Ri j < rval. Depending on which expression(s) that are used (determined by parameter ival) the relevant parameters, R_0 and B or R_0 and N, should be given in the control data file.

6. RMC - why use it?

There are numerous methods of structural modelling, from the simplest 'hand built' or 'hand drawn' models to conventional MC or MD simulations. However RMC has several advantages.

1. RMC uses all the available structural data, not just particular features, in a quantitative rather than qualitative manner. Many models that use particular features, for example peak positions and coordination numbers from radial distribution functions, can be misleading.
2. RMC is potential independent. If a potential exists that, when used in an MC or MD simulation, also produces quantitative agreement with experimental structure factors, then this is obviously equally as good as RMC (and one would hope that the results were similar!). However few potentials provide such quantitative agreement, and in some cases it has not yet proved possible to produce potentials that provide qualitative agreement. Also most simulations are compared to experiment at $g(r)$ stage because the configurations are too small to allow transformation to A(Q). For the reasons discussed above it is important for good structural modelling that comparison be made at the A(Q) stage.
3. Because RMC models a three dimensional structure $g^C(r)$ and $A^C(Q)$ must correspond to a possible physical structure; the model is subject to the simple but powerful constraints of fixed density and excluded volume (minimum atomic sizes). However $g^E(r)$ derived by conventional methods may contain errors which

mean that it could not correspond to a possible physical structure; that is, it is internally inconsistent. In the case of multicomponent, systems there is no requirement that the partial radial distribution functions derived by conventional methods are consistent with one another, while in RMC they must be consistent and physically possible. This constraint improves the separation of partials in cases where the separation matrix is poorly conditioned. It also means that some information on partials can be obtained for underdetermined cases, where none can be obtained by conventional methods.

4. Different types of data, for example neutron and X-ray diffraction, can be modelled. The different data sets can have different Q ranges, spacings, resolutions etc. It is also easy to include additional constraints on the structure; these could be from other experimental information (e.g. NMR) or could be some other knowledge of, or assumptions about, the system (e.g. chemical bonding ideas).

5. RMC is easily adapted to different physical problems.

7. Uniqueness

The three dimensional structure produced by RMC is not unique, it is simply a model that is consistent with the data and any additional constraints. Other methods that produce structures which are equally consistent with the data are equally valid and there is no way of determining which is 'correct' in the absence of any *additional information*. One possible disadvantage of RMC is that it tends to produce the most disordered structure that is consistent with the data and constraints, that is the configurational entropy is maximised. However this is counteracted by the ability to include additional constraints, which means that additional ordering can be imposed and a range of consistent structures investigated; those that are found to be inconsistent can then be discounted.

In the special case of a system for which the interatomic potential is purely pairwise additive there is a theoretical justification for the determination of the three dimensional structure from a one dimensional g(r) or A(Q) [11]. Given that the potential uniquely determines the structure

$$\phi(r) \Rightarrow g^{(2)}(r_1, r_2), g^{(3)}(r_1, r_2, r_3), g^{(4)}(r_1, r_2, r_3, r_4)\ldots \tag{24}$$

where $g^{(2)}(r_1, r_2) = g(r)$ and $g^{(n)}(\ldots)$ are the *n*-body correlation functions, then for a pairwise additive potential there is a functional relationship between $\phi(r)$ and g(r) such that

$$g^{(2)}(r_1, r_2) \Rightarrow \phi(r) \tag{25}$$

that is that g(r) uniquely determines ϕ(r). (This is not to say that we can write down the relationship, but merely that one exists.) If g(r) determines ϕ(r) and ϕ(r) determines the structure, then g(r) determines the structure

$$g(r) \Rightarrow g^{(3)}(r_1, r_2, r_3), g^{(4)}(r_1, r_2, r_3, r_4)\ldots \tag{26}$$

Theoretical tests have shown that in cases where the potential is purely pairwise additive the RMC method works satisfactorily [21].

The potentials in real systems are never purely pairwise additive (though such potentials are used in the majority of MC and MD simulations). However the above result does indicate that a precisely measured g(r) or A(Q) does contain a great deal of information about the three dimensional structure. RMC is one possible way of attempting to extract this information. Where there are significant three-body terms in the potential then constraints may be used to take account of them. In the case of molecular liquids, for example, molecules can be included explicitly in the model.

8. Applications of RMC

There are now numerous different applications of RMC modelling. Here they are summarised, with references, both by subject and technique.

1. RMC method [4,8,11,17,18,21,27,28,33-35,37].

2. Neutron diffraction [4-6,8-16, 18-20,22-29,31-34,36-38,40-43,46,47].

3. X-ray diffraction [8,9,14,24,28,33,39,44,45].

4. EXAFS [17].

5. Use of constraints [19,22,28,33,34,49].

6. Combination of experimental techniques [8,14,24,28,33,34,36].

7. Elemental liquids: condensed inert, gases [4,23,32], liquid metals [20,32], liquid semiconductors [32], molecular liquids [9,32].

8. Binary liquids: liquid metal alloys [19,42,43], liquid semiconductors [6], molten salts [5,151].

9. Aqueous solutions [10].

10. Covalent glasses: silicates [141, Ag^+ fast ion conducting glasses [25,31,36], amorphous carbon [24].

11. Metallic glasses [38-41,44-46].

12. Magnetism in metallic glasses [22].

13. Structural disorder in crystals: Ag^+ fast ion conductors [12,13,26,29], C_{70} [30].

14. Polymers [47].

9. RMC - simulation details

9.1 Configuration size and shape

When starting from the initial configuration $g(r)$ must be calculated. This involves a summation of order N^2. However for each particle move it is only necessary to calculate the change in $g(r)$ corresponding to the moved particle, which is a summation of order N. This is the same in MMC, but not in Molecular Dynamics (MD) where all moves are of order N^2 (unless the potential is truncated). For this reason MC simulations may involve much larger configurations than are used in MD. Generally we use $N > 1500$, and have used $N \approx 30000$. The size of simulation is important when modelling $A^E(Q)$ since $g^C(r)$ may only be calculated up to $r = L/2$. In order to be able to transform $g^C(r)$ directly to $A^C(Q)$ we require that $g^C (r > L/2) = 1$. Any significant deviations from this, either due to long range correlations or statistical fluctuations, will cause truncation ripples at low Q in $A^C(Q)$. Size is also relevant when modelling $g^E(r)$, because this determines the statistical fluctuations in $g^C(r)$ and hence the effective value of $\sigma(r)$.

If there are long range correlations, such as in crystalline materials, it is not possible to make the box large enough for $g^C(r > L/2) = 1$ to be a good approximation. However, there is a way around this problem. The effect of the finite box size on the calculated structure factor is that, in the sine Fourier transform that calculates $F^C(Q)$, the argument $g(r) - 1$ is multiplied by a step function that is unity for $r < L/2$ and zero for $r > L/2$. Using the convolution theorem for Fourier transforms it is possible to show that this is equivalent to convolution of $F^C(Q)$ to give

$$\frac{1}{\pi}\int_{-\infty}^{\infty}F^C(Q')\frac{\sin L/2(Q-Q')}{Q-Q'}dQ'$$

$$=\frac{1}{\pi}\int_{0}^{\infty}F^C(Q')\left[\frac{\sin L/2(Q-Q')}{Q-Q'}-\frac{\sin L/2(Q+Q')}{Q+Q'}\right]dQ' \qquad (27)$$

In order to compare with the experimental data we should therefore perform the same convolution before using it as input to the RMC program. A program CONVOL is available for this purpose. However, before this operation is performed it is essential to know any offset in the experimental data. Because this offset is normally determined by background parameter(s) that are refined during the RMC simulation, it is in general recommended to apply the internal convolution option. A further advantage of doing so is that the convolution is then automatically performed with the appropriate value of L, the value used during the simulation.

An alternative approach for crystals is to model the radial distribution function using the program MCGR, and then fit this. The $g(r)$ can be modelled out to a large enough r value, typically 150-200 Å, so that the experimental Q resolution is matched and there is no truncation in the r to Q Fourier transform. Only the lower r part of this, i.e. $r < L/2$, is fitted by RMC.

9.2 Closest approach distance of two atoms

For perfect data the distances of closest approach of pairs of atoms are determined by the low r cut-off in $g^E(r)$. However for imperfect data, particularly when $A^E(Q)$ is significantly truncated at the maximum Q value, the closest approach may not be well defined. For this reason it is usually sensible to specify allowed distances of closest approach, in other words to define an excluded volume. This also saves considerable time since moves which would result in atoms being too close together can be rejected before calculation of the change in $g^C(r)$. For good data the specified closest approaches may be somewhat lower than realistic values but for poor data they need to be more carefully chosen. If the values are too large then this is usually apparent because the resulting $g^C(r)$ has a sharp cut-off instead of decreasing more gradually to zero. If they are too low $g^C(r)$ may have a sharp spike in the low r region.

While this is a very simple constraint on the structure it is also very powerful, since the imposition of both an excluded volume and a fixed density restricts possible configurations. One could also view it as the imposition of a hard sphere repulsive potential. In the case of, for example, a two component system where the hard sphere radii are sufficiently different this constraint allows one to obtain some information on all three partial radial distribution functions or structure factors from one or two total structure factors, while three are required for a conventional solution. This is valuable in cases where suitable isotopes are not available for neutron diffraction. Since the resulting structure is then dependent on the choice of radii one should, if possible,make a choice based on other experimental information rather than treating them as free parameters.

9.3 Maximum size of random move

The maximum size, δ, of the random move determines the ratio of accepted to rejected moves, but also determines the amount that the structure may change with each move. If δ is too small then nearly all moves will be accepted but the structure will change little, while if it is too large then few moves will be accepted and the average structural change will also be small. If we attempt to choose δ such that the ratio of accepted to rejected moves is approximately one, as is often done in MMC, this usually leads to a value $\delta < 0.1\text{Å}$. The average structural change per move is usually maximised for $0.1 < \delta < 0.3\text{A}$ with an acceptance/rejection ratio approximately 0.5, so this range is normally used.

When starting from a structure that is significantly different from the 'real' structure it is possible that certain atoms become 'trapped' in some local arrangement. This is a local minimum in the minimisation procedure, rather than the required global minimum. One way around this problem is to run the simulation

for a while with a large value of δ, for instance up to 10Å. While hardly any moves will be accepted those that are may be sufficient to get the configuration out of the local minimum. (The terms 'local' and 'global' are used here in an unconventional manner; 'global' refers to any minimum that satisfies the required fitting criterion and 'local' to any minimum that does not. We specifically do not wish to attain the conventional global minimum, the single structure that is closest to the data, since the data contain errors.)

If the packing fraction (ratio of excluded volume to total volume) is high, as with many metallic glasses, then it is necessary to use small moves so that a sufficient number are accepted. The convergence of the RMC procedure may then be very slow. One way to 'speed up' the process is to artificially decrease the cut-offs for some time, i.e. to decrease the packing fraction, and then to increase them again later when a better fit to the data has been achieved.

Another issue with relation to dense packing regards possible chemical ordering. In order to speed up convergence for such cases it is also possible to directly swap the positions of two atoms of different types. The appropriate fraction of such moves depends on the actual system. In any case, the standard displacive moves are also needed to obtain a good fit, so swapfrac should be < 1.

See subsection 9.8 for a discussion of move sizes in relation to the use of coordination constraints.

9.4 r spacing and Q range

When modelling $A^E(Q)$ the minimum r spacing is determined by the maximum Q value, Q_m The real space resolution is then $2\pi/Q_m$; one requires approximately five points over this range so an r spacing of $2\pi/(5Q_m)$ is appropriate. For simple liquids where structure in $A^E(Q)$ extends out to $Q \approx 10\text{Å}^{-1}$ this makes a spacing of approximately 0.1Å suitable, whereas for molecular liquids or glasses with structure out to $Q \sim 40\text{Å}^{-1}$ a spacing of 0.025 - 0.05Å is suitable. However it should be noted that decreasing the r spacing for a fixed number of particles increases the statistical error in $g^C(r_j)$, so it may be necessary to increase the model size. Having a large number of r points also increases the transform time if structure factors are being modelled, and increases the size of the coefficients file if EXAFS data are being modelled. Some compromise may be necessary if all of these factors are taken into account.

The minimum Q value that can be modelled is given by $Q_{min} = 2\pi/L$ If you try to fit to smaller Q values then the effects are unpredictable. For example, if a much smaller Q_{min} is used then this can lead to distinct density fluctuations of period $2\pi/L$ in the configuration.
Note that weak fluctuations of this period can be seen in many simulations, not just in RMC.

9.5 experimental error: σ

The RMC algorithm assumes that we have only statistical errors. In practice this is not true, but the whole procedure is not thereby invalidated. A three dimensional structure that is consistent with the experimental data within some measure of the error can still be produced, though this measure is now less well defined.

A real experimental structure factor $A^E(Q)$ will contain both statistical and systematic errors. While one might expect statistical errors to be small where $A^E(Q)$ is large, and vice-versa, in practice the requirement to perform container and background corrections in many experiments means that statistical errors are often quite uniformly distributed. In many X-ray experiments counting times are chosen to deliberately produce a uniform distribution. Since we often have no knowledge of the likely distribution of systematic errors it is usually simplest to assume a constant value of σ at all Q, though σ may differ between different data sets. However there have been cases in which large values of σ have been used in particular Q ranges where it is known that there were errors in the data. By setting $\sigma(Q)$ at an extremely large value these data points can effectively be ignored.

When $A^E(Q)$ is transformed to $g^E(r)$ the errors are redistributed. There are also statistical errors inherent in $g^C(r)$. Comparison at the g(r) stage will therefore not produce precisely the same result as comparison at the A(Q) stage. This is of course more exaggerated when additional errors are introduced in $g^E(r)$ by truncation of $A^E(Q)$. It is also worth noting that certain features of $A^E(Q)$, in particular 'prepeaks' or 'first sharp diffraction peaks' in some glasses and liquids at $Q \approx 1\text{Å}^{-1}$, correspond to small, long period modulations

of $g^E(r)$. It is possible that such real space structure can be 'ignored' to a great extent if the modulation amplitude is comparable to the chosen value of $\sigma(r)$. For these reasons it is strongly recommended that RMC is used for modelling $A^E(Q)$ wherever possible. However it is also possible that in cases where there are well defined peaks in $g(r)$, for instance covalent bonding, and correspondingly oscillations in $A^E(Q)$ out to large Q, the high Q structure factor is not fitted as well as the low Q part. The peaks in $g^C(r)$ will then be lower and broader. To overcome this both $A^E(Q)$ and $g^E(r)$ (preferably obtained by an inverse method such as the MCGR program) can be fitted simultaneously.

From the above discussion it is clear that the precise value of a is not known in any particular case; it may therefore be considered as a parameter of the simulation.

If we make the analogy with MMC that $\chi^2 = U/kT$ then a corresponds to kT. Under normal circumstances we would use a value of a that was approximately 1% of the amplitude of the input data (a typical value of experimental error). However, if it is believed that the simulation has run into a local minimum then, as is common procedure in other simulations, we would increase the value of σ (analogous to increasing the temperature). After running the simulation for a while σ would then be decreased to its original value. Alternatively if we deliberately wished to find a local minimum closest to a particular starting configuration we can effectively set $\sigma = 0$ by only accepting moves which decrease χ^2.

We have generally found for disordered structures that the global minimum in χ^2 is relatively broad and little manipulation of σ or d is required to reach it. However for more ordered structures, such as crystals, this is not the case and global minimisation of χ^2 can only be achieved by simulated annealing with σ as the control parameter.

9.6 Renormalisation and background

Experimental data will normally contain small normalisation errors in the form of multiplicative and additive constants. It is possible to take account of such errors within the RMC algorithm. This can be particularly important when dealing with isotopic substitution neutron diffraction data when the relative normalisation of different structure factors must be correct. It is also useful for correcting for incoherent inelastic scattering for hydrogenous materials. During the RMC process, the real experimental total structure factors, $F^E(Q)$, can be considered as simple transformations of the structure factors actually measured, $F^E_m(Q)$ since the latter may contain systematic errors:

$$F^E(Q) = \left(F^E_m(Q) - \sum_{i=0}^{n_B} a_i Q^i \right) / b \qquad (28)$$

where n_B is the degree of a background polynomial. The expression that is actually minimised in the simulation procedure is then accordingly

$$\chi^2 = \sum_{i=1}^{m} \left(bF^C(Q_i) + \sum_{j=0}^{n_B} a_j Q^j - F^E(Q_i) \right)^2 / \sigma(Q_i)^2 \qquad (29)$$

Optimisation of any combination of the parameters is possible; it is recommended, however, that the renormalisation factor (parameter b) be refined only when the best possible fit without changing it could be achieved. If the required value differs significantly from 1 then the experimental data should obviously be checked. The values of the above parameters, together with the χ^2 value, provide a sensitive test of data quality, particularly in terms of systematic errors.

9.7 Definitions of structure factors etc.

For the purposes of RMCA the input data, or structure factors, are defined in the following way. Data which are not defined in this way should be modified accordingly.

$$G(r) = \sum_{\alpha=1}^{n} \sum_{\beta=\alpha}^{n} \gamma_{\alpha\beta} \left(g_{\alpha\beta}(r) - 1 \right) \tag{30}$$

Where $\gamma_{\alpha\beta}$ are coefficients which are specified. G(r) is the function output by the program MCGR, so this can be used directly as input to RMCA if required. If you wish to model a partial radial distribution function, e.g. $g_{12}(r)$, then all coefficients except γ_{12} must be set to zero and 1 must be subtracted from the data. It is not possible to subtract an offset value from G(r) within the program since it must by definition tend to zero at large r.

$$S(Q) = \sum_{\alpha=1}^{n} \sum_{\beta=1}^{n} \gamma_{\alpha\beta} \left(A_{\alpha\beta}(Q) - 1 \right) \tag{31}$$

Where $\gamma_{\alpha\beta}$ are coefficients which are specified. Partial structure factors can be modelled in the same way as partial radial distribution functions. Note that G(r) is the direct transform of S(Q).

$$F(Q) = \sum_{\alpha=1}^{n} \sum_{\beta=1}^{n} \gamma_{\alpha\beta}(Q) \left(A_{\alpha\beta}(Q) - 1 \right) \tag{32}$$

Where $\gamma_{\alpha\beta}(Q)$ are Q dependent coefficients which are given in the same file as F(Q). Note that the direct transform of F(Q) *is not* G(r), because of the Q dependence of the coefficients. Normally for X-ray data we define

$$\gamma_{\alpha\beta}(Q) = \frac{c_\alpha c_\beta f_\alpha(Q) f_\beta(Q)}{\sum_\alpha c_\alpha f_\alpha(Q)^2} \tag{33}$$

Where $f_\alpha(Q)$ is the form factor for species a. F(Q) then tends to a constant value at high Q and the offset and renormalisation options can be used.

For EXAFS data the coefficients are both Q and r dependent, and the structure factor is defined as

$$E_\alpha(Q) = \sum_\beta 4\pi \rho \int r^2 \left(g_{\alpha\beta}(r) - 1 \right) f_\beta(Q,r) \, dr \tag{30}$$

The values of $f_\beta(Q,r)$ are given in a separate file from the data.

9.8 Use of coordination constraints

Coordination constraints are one of the most valuable and instructive 'tools' in the RMCA program. A description of how they may be used is best given by some examples.

9.8.1 Molecular liquids.

A model of water can be constructed as follows. The H-O-H molecule can be contained in a sphere of radius 1.6 Å centred on the O atom; the inter-molecular H-H distances are then larger than the intra-molecular distance. We start with a random arrangement of O atoms at a density of 0.03333 atoms Å^{-3} and run a hard sphere simulation (RMC with no data), with a constraint on the closest approach of 3.2 Å, until no atoms are too close together. Two H atoms are then added to each O atom with relative positions (x, y) and $(-x, y)$ where $x = 0.7572$ Å and $y = 0.5868$ Å. The resulting configuration is now a set of aligned H-O-H molecules at a density of 0.09999 atoms Å^{-3}. The hard sphere simulation is run again, with the constraints that each O is coordinated to two H between 0.0 and 1.0 Å and each O is coordinated to no H between 1.0 Å and 1.05 Å. This destroys the molecular alignment, the second constraint ensuring that the inter- and intra- molecular H-H distances are distinctly separate. The constraints are maintained and the data fitted. The molecules will remain 'bonded' but are flexible; the H-O-H bond angle should be determined by the data. If a bond angle constraint is required then this can be done by requiring each H to have one H neighbour (that in the

14

molecule) within appropriate distances, for instance 1.1 Å and 1.2 Å. Because the constraint on the O-H bond length is severe only small move sizes (of order 0.05Å) can be used and convergence is slow.

9.8.2 Network glasses.

Si in silicate glasses are coordinated to four O. The SiO_4 tetrahedra share O at each comer (bridging oxygens) with other tetrahedra. In pure SiO_2 all oxygens are bridging; when alkali oxide is added some comer O are no longer shared and these are known as non-bridging. Tetrahedra with n bridging O are known as Q_n species, and their relative proportions can be found from MAS NMR data.

To model $(K_2O)_{0.15}(SiO_2)_{0.85}$ glass, for example, we make an initial configuration of random Si at a density of 0.0174 atoms A'. Two coordination constraints are applied, corresponding to 16% 3-fold coordination and 84% 4-fold coordination between 2.9Å and 3.4Å, and a hard sphere simulation is run until the constraints are satisfied. In the later stages of this process it may be necessary to move over- and under-coordinated atoms around the configuration by hand, otherwise the process may take a very long time. O atoms are added at the mid-point of each Si-Si bond, and single O atoms are added close (e.g. at a distance of 1.0 Å) to all Si atoms which are then coordinated to less than four O, unless 100% Si-O 4-fold coordination within 2.0 Å is obtained. The density is now 0.0553 atoms $Å^{-3}$. Another hard sphere simulation is run until all atoms obey the correct cut-off constraints. K atoms are added at random, the density increasing to 0.0614 atoms $Å^{-3}$, and a final hard sphere simulation is run until they satisfy the cut-off constraints. The resulting configuration now has the correct topology. The data can then be fitted with the single coordination constraint of 100% Si-O 4-fold coordination within 2 Å. Because this constraint ensures that the toplogy of the whole network remains unchanged it is extremely severe, so only small move sizes (of order 0.05Å) can be used and convergence will be slow.

9.9 Metropolis Monte Carlo

The RMCA program may also be used for conventional MMC simulation with either hard spheres or a potential. Since the imposition of closest approach distances for atoms is equivalent to having hard spheres, for a hard sphere simulation these distances should be chosen as appropriate and no experimental data specified, i.e. the number of experimental data sets is zero. In most cases it would be recommended that a suitable hard sphere simulation be run before any attempt is made to fit to the data.

When a potential is used it is defined in a table at the same r interval as used to define g(r). This makes the calculation of the energy very fast, but is not a suitable method for a 'proper' MMC simulation. In this case it is only intended that the potential be used to produce an initial configuration, or as an additional constraint for the RMC procedure.

9.10 Efficient use of RMC

The time taken by the `RMCA` program depends on the number of atoms in the configuration, the number of r points used and the number of data points and data sets. For any particular application the time required can vary from hours to months, so it is sensible to have a strategy which will reduce this as much as possible. The general approach should be as follows.

1. Use parameters appropriate for the problem, i.e. do not use enormous configurations (though always more than 1000 atoms) etc.

2. Create an initial configuration which satisfies all constraints. If only closest approach constraints are being used then this may be possible using the `MOVEOUT` program. Otherwise `RMCA` must be run with no experimental data. If coordination constraints are being used then this first step may take a long time.

3. If using diffraction data only then initially fit to G(r) or $g_{\alpha\beta}(r)$ (preferably obtained using `MCGR`). This is better than fitting directly to the structure factor as the transform from r to Q is expensive. Fit to all sets of

diffraction data available. Fitting to single sets and then adding others does not generally save time, unless the data sets contain very similar information anyway. This can be assessed using the program PARTIALS.

4. Fit to a subset of Q points in the structure factors, e.g. at 0.1 Å^{-1} intervals, unless you are fitting to diffraction data from crystals when the resolution provided by the experimental Q spacing is required.

5. Add in EXAFS data if it is being used. Since EXAFS only provides information on short range order it is generally safe to obtain a good fit to diffraction data first. However this will not always be the case and it may then be necessary to relax the fit to the diffraction data before the EXAFS data can be fitted satisfactorily.

6. Finally fit to all the experimental data points, though do not use more than are justified by the experimental resolution.

10. The RMCA program

10.1 Running the program

The current version of RMCA (version 3) is written as a general purpose program for modelling multi-component systems using experimental diffraction data as a constraint. When RMCA is run it must be supplied with a name used for all its data files. This is usually done on the command line, depending on the computer on which the program is installed. For example

<p align="center">rmca cscl</p>

would run the RMCA program with files called cscl with various extensions. This name is given to a number of files used by the program. The general program parameters are supplied in name.dat. The first time that the program is run there must be a file name.cfg containing the positions, or configuration, of the atoms. It is only necessary to calculate g(r) once at the beginning of the run, thereafter only the change in g(r) needs to be calculated. To facilitate this the histograms representing g(r), as well as a binary representation of the atomic coordinates, are written out to a file name.his when the program terminates; the new configuration is also written separately to name.cfg. On the next run the histograms and the atomic coordinates will be read from name.his, rather than from name.cfg, unless any parameters have been changed which require the histograms to be recalculated. Saved configurations will be written to name.sav and the output results will be written to name.out. Information on the status of the program is written to name.log. An example log file is shown in Appendix I.

The Windows version is simply started by clicking the program icon, or it may be started from the RMC menu using the WinNFLP program. An Open file window appears where the user can choose the relevant name.dat file and then the simulation will start. During the run menus can be accessed to save current data promptly or to exit the program with or without saving. Plotting of the current fitted patterns and the progress of χ2 can also be turned on/off. A record of the last work directory will be saved to a file rmca.ini in the directory where the exe file is stored so that easy continuation is possible next time the program is run.

10.2 The configuration file

The configuration is normally cubic, but in general any parallelepiped can be used. If a crystalline system is being modelled then obviously the box dimensions must be chosen to accomodate an integral number of unit cells in each direction, with the numbers being chosen to make the dimensions as equal as possible. The atomic coordinates are defined in terms of the box coordinates and are normalised to limits of ±1.0. The box coordinates are then defined in terms of the laboratory coordinates by the matrix given in the top of the configuration file. The values in this matrix are strictly only required to be relative; their absolute values are recalculated from the density every time the program is run and these are then written into the new

configuration file. However some programs for analysis of configurations do not check the density (and may produce erroneous results) so it is sensible to always use absolute values. For a cubic box the diagonal terms in the matrix are equal to half the box length (L/2) in Å.

The numbers of different types of atoms and the order of their coordinates are given in the top of the configuration file. The atomic coordinates then follow in a single list. The choice of a format in which the system is not explicitly identified in the configuration file (other than by the title) is made deliberately; configurations can then be easily modified or used as starting configurations for different systems with a minimal amount of editing. The same format of configuration file is also used for molecular systems, where both coordinates and Euler angles define the molecules. For this reason atoms are defined as molecules with a single atomic site. An example configuration file is shown in Appendix II.

10.3 The control data file

The RMCA data file `name.dat` can be thought of as being divided into various sections. The description that follows is given section by section. An example data file is shown in Appendix III.

10.3.1 General parameters.

title	character*80	A title for the run.
rho	real	Number density of system in atoms Å^{-3}.
rcut	real array	The closest allowed approach of two particles. There should be a value for each partial g(r). There are thus $n(n + 1)/2$ values for a n-component system. As in all other parts of the program the partials are given in the order 1-1, 1-2, ..., 1-n, 2-2, ..., 2-n, ... n-n. See section 9.2 for a description of the use of this parameter.
delta	real array	The maximum move for each type of particle. Recommended values are in the range 0.05-0.5. A value of zero is allowable, in which case the program will not attempt to move those particles. See section 9.3 for a description of the use of this parameter.
nswap, swapfrac	integer,real	The number of kinds of swap moves and the fraction of generated moves that will be of the swap kind.
swaptypes	2*integer	The first and second particle types to swap. One line for each nswap kind.
dr	real	The spacing for calculating g(r). See section 9.4 for a description of the use of this parameter.
moveout	logical	This should be set .true. if, and only if, there are some particles that do not satisfy the cut-off restrictions and you want to preferentially choose these to be moved. 10% of the time one of these particles will be chosen, and 10% of the time any particle will be chosen.
ncoll	integer	The number of configurations to collect after convergence. The program will stop when ncoll configurations have been collected.
iprint,iplot	2*integer	Determines how often a summary will be written to the standard output and how often a plot of the current results will be updated. It will be written after every iprint moves generated, except that this will only occur when a move is accepted. iprint should not be too small or otherwise time will be wasted in continually writing to the .log file and this will become very

large.

If iplot is 0 plotting is initially turned off. In the Windows version it can be turned on and controlled during the run using the Plot menu items and the mouse (left-click and drag to select region, right-click to reset).

timelim, timesav	2*real	The time the program should run for, in minutes.

The interval at which the results should be saved to the output files (they are always saved when the program ends anyway). If the program is left running for a very long time it is best to save the results every now and then, perhaps once an hour. The results should not be saved too often as this simply wastes time writing large files.

10.3.2 Parameters referring to all data sets.

nexpt	integer array	Four values, being the number of experimental data sets of different types. The first is the number of G(r) constraints, the second the number of S(Q)'s, the third the number of F(Q)'s, (for the purpose of this program the difference between S(Q) and F(Q) is that the former has constant coefficients of partial structure factors whereas the latter has Q dependent coefficients as may be the case with x-ray data), and the fourth the number of EXAFS data sets. Definitions of these quantities in terms of partial structure factors and radial distribution functions are given in section 9.7.

10.3.3 Parameters for each G(r) constraint.

file	character*80	File containing the experimental data.
nxl,nx2	integers	The indices of the first and last data point to be used. If nx2 is larger than the number of data points the last data point to be used will be the last data point given. If the r value corresponding to nx2 is larger than L/2 then nx2 will be reduced appropriately
coeff	real array	The coefficients of each of the partial g(r)'s.
isig,sigma	integer,real	isig - code for determining the experimental standard errors. If isig = 1 standard errors are read from an extra column in the experimental data file. If isig = 2 standard errors are estimated as the square root of the experimental data. If isig = 3 standard errors are constant = 1; sigma - standard errors obtained as above are scaled by this value to yield the $\sigma(Qi)$ in eq. (5).
renorm	logical	.true. if data renormalisation is required.
beta	real	Initial normalisation factor.
offset	logical	.true. if refining the polynomial background.
nbckg	integer	no. of terms in the polynomial background.
bcoeff	nbckg*real	the initial coefficients of the polynomial background.

10.3.4 Parameters for each S(Q) constraint.

file	character*80	File containing the experimental data.
nx1,nx2	integers	The indices of the first and last data points in the experimental

		data file. If `nx2` is larger than the number of data points the last data point used will be the last data point given.
coeff	real array	The coefficients of each of the partial structure factors.
isig,sigma	integer,real	isig - code for determining the experimental standard errors. If isig = 1 standard errors are read from an extra column in the experimental data file. If isig = 2 standard errors are estimated as the square root of the experimental data. If isig = 3 standard errors are constant = 1; sigma - standard errors obtained as above are scaled by this value to yield the $\sigma(Q_i)$ in eq. (5).
convol	logical	`.true.` if data convolution is required, see 9.1.
renorm	logical	`.true.` if data renormalisation is required.
beta	real	Initial normalisation factor.
offset	logical	`.true.` if refining the polynomial background.
nbckg	integer	no. of terms in the polynomial background.
bcoeff	nbckg*real	the initial coefficients of the polynomial background.
magnetic	logical	`.true.` if a paramagnetic form factor found in the structure factor file, in the same way as for X-ray form factors, will be added to the calculated structure factor before fitting to the data.

10.3.5 Parameters for each F(Q) constraint.

file	character*80	File containing the experimental data and the coefficients of the partial structure factors.
nx1,nx2	integers	The indices of the first and last data point to be used. If `nx2` is larger than the number of data points the last data point to be used will be the last data point given.
isig,sigma	integer,real	isig - code for determining the experimental standard errors. If isig = 1 standard errors are read from an extra column in the experimental data file. If isig = 2 standard errors are estimated as the square root of the experimental data. If isig = 3 standard errors are constant = 1; sigma - standard errors obtained as above are scaled by this value to yield the $\sigma(Q_i)$ in eq. (5).
convol	logical	`.true.` if data convolution is required, see 9.1.
renorm	logical	`.true.` if data renormalisation is required.
beta	real	Initial normalisation factor.
offset	logical	`.true.` if refining the polynomial background.
nbckg	integer	no. of terms in the polynomial background.
bcoeff	nbckg*real	the initial coefficients of the polynomial background.
compton	logical	`.true.` if Compton scattering, found in the structure factor file in a column following the *Q* dependent coefficients, will be added to the calculated structure factor before fitting to the data.

10.3.6 Parameters for afl EXAFS data (only present if there is some).

rmax	real	Maximum *r* value to be used.
weight	integer	Data are to be weighted by Q^{weight}. The experimental data should be supplied unweighted, as weighting is done within the program. Data and fits written in the `.out` file will have the required *Q* weighting.

10.3.7 Parameters for each EXAFS constraint.

file	character*80	File containing the experimental data.

nx1,nx2	integers	The indices of the first and last data point to be used. If nx2 is larger than the number of data points the last data point to be used will be the last data point given. When EXAFS data is modelled together with diffraction data the values of nx1 and nx2 *must* correspond to *Q* values which are a subset of the *same Q* values as in the data files for the diffraction data, and similarly in the file of EXAFS coefficients, as the *Q* values will be taken from the diffraction, data files.
nx3,nx4	integers	The indices of the first and last data points to be used in calculation of χ^2 for the EXAFS data. nx3 and nx4 must lie within the range of nx1 and nx2. These additional parameters are given for EXAFS data so that different ranges of the EXAFS data can be modelled without having to recalculate the EXAFS coefficients.
edge	integer	Index number of the particle type corresponding to the absorption edge being used.
file	character*80	File containing the EXAFS coefficients.
isig,sigma	integer,real	isig - code for determining the experimental standard errors. If isig = 1 standard errors are read from an extra column in the experimental data file. If isig = 2 standard errors are estimated as the square root of the experimental data. If isig = 3 standard errors are constant = 1; sigma - standard errors obtained as above are scaled by this value to yield the $\sigma(Qi)$ in equation (5).
convol	logical	.true. if data convolution is required, see 9.1.
renorm	logical	.true. if data renormalisation is required. It is not recommended that this option be used with EXAFS data since renormalisation by a negative number, corresponding to a phase shift of π, may occur.
beta	real	Initial normalisation factor.
offset	logical	.true. if refining the polynomial background.
nbckg	integer	no. of terms in the polynomial background.
bcoeff	nbckg*real	the initial coefficients of the polynomial background.

10.3.8 Parameters for coordination constraints.

ncoord	integer	The number of coordination constraints.

10.3.9 Parameters for each coordination constraint (one line per constraint).

typec	integer	The type of the central particle.
typen	integer	The type of the neighbour particles.
rcoord	2*real	The two distances between which to calculate the coordination number.
coordno	real	The desired coordination number.
coordfrac	real	The fraction of the central particles desired to have this coordination number.
sigmac	real	Effectively a parameter weighting this constraint relative to others and the fit to the data.

10.3.10 Parameters for average coordination constraints.

ncoord	integer	The number of average coordination constraints.

10.3.11 Parameters for each average coordination constraint (one line per constraint)..

typec	integer	The type of the central particle.
typen	integer	The type of the neighbour particles. Note that for the average coordination number constraints, unlike the constraints on coordination number distribution, there is no difference in the constraint obtained by swapping neighbour and central particles except in the value of coordination number required, as the constraint is simply obtained by integration of $g_{ij}(r)$ and $g_{ij}(r) = g_{ji}(r)$.
rcoord	2*real	The two distances between which to calculate the coordination number.
coordno	real	The desired average coordination number.
sigmac	real	Effectively a parameter weighting this constraint relative to others and the fit to the data.

10.3.12 Parameters for bond valence sum constraints.

nval	integer	The number of bond valence sum constraints.

10.3.13 Parameters for each BVS constraint (one line per constraint)..

typec	integer	The type of the central particle.
nvaltypen	integer	No. of neighbour types.
rval	real	The maximum bond distance for which to calculate the bond valence sum.
valreq	real	The required average of the bond valence sum.
sigmaval	real	Effectively a parameter weighting this constraint relative to others and the fit to the data.
valrms	real	The required R.M.S. of the bond valence sum.
sigmavalrms	real	Effectively a parameter weighting this constraint relative to others and the fit to the data.

10.3.13.1 Parameters for each neighbour type (one line per type).

valtypen	integer	The type of the neighbour particles.
ival	integer	Code to select bond valence expression, ival =1 → power law, ival = 2 → exponential.
r0val, bval	2*real	Parameters for the ival expression, see 5.

10.3.14 Parameters for angular triplet constraints.

ntrip	integer	The number of angular triplet constraints.

10.3.15 Parameters for each angular triplet constrait (oneline per constraint).

typea,typeb, typec	3*integer	The types of the particles in the triplets A-B-C.
tripr1, tripr2	2*real	Minimum and maximum distance between two atoms considered as part of a triplet, i.e. only triplets with tripr1 $< r_{AB} <$ tripr2 and tripr1 $< r_{BC} <$ tripr2 are used.
tripang	real	The desired triplet angle (in degrees).
tripsig	real	Effectively a parameter weighting this constraint relative to others and the fit to the data.

10.3.16 Parameters concerning a potential.

usepot	logical	`.true.` if using a potential. The remaining parameters in this section are only present if we are.
temp	real	The absolute temperature.
eunits	real	The energy units the potential is supplied in (in other words the conversion factor to Joules).
weight	real	The weighting factor for the potential. To use the program. for a conventional Metropolis Monte Carlo simulation this should be 1.
file	character*80	The data file containing the potential tabulated at intervals of d*r*.

11. The experimental data files

The files containing the experimental data are in the DATA format as defined for the NDP series of programs. The number of points is given on the first line, the second line contains a title or other information, and the subsequent lines contain the *Q,* S*(Q)* or *r*, G*(r)* values in two columns. If **magnetic** is .true., in the case of S(Q) data, then a third column contains a paramagnetic form factor which will be added to the calculated structure factor before the data are fitted. An example file is shown in Appendix III.

If the coefficients of the partial structure factors are to be included in the files, i.e. for F*(Q)* data as defined in subsection 9.7, then they come after the F*(Q)* values in the order 11,12,...1n,22,23,...2n,...nn. For X-ray data files in this format can be produced from files in the DATA format using the program XCOEFF. This has the option to normalise the form factors in three different ways, depending on how the original data have been defined. If **compton** is .true. then a final column contains Compton scattering which will be added to the calculated structure factor before the data are fitted. An example file is shown in Appendix IV.

Note that if more than one set of diffraction data is supplied they must all be defined at the same *Q*, or *r*, points. You only need to use a subset of these points for fitting so it is possible to use data sets that cover different ranges provided they are defined (for instance set to zero) at the points where data are not available. In the examples given in Appendix IV the second data set covers the range 0.3 to 15.9 at 157 points. The first data set only covers the range 0.6 to 9.8, so it is defined to be zero at the other points. Points 4 to 96 are used for fitting to the first file (`zncl235.fq` in Appendix III) and points 1 to 157 for fitting to the second (`zncl2x.fq` in Appendix IV). The NDP program REBIN can be used to produce data sets satisfying this requirement.

If both EXAFS and diffraction data are supplied then the EXAFS data must be defined at a subset of the same *Q* values as the diffraction data, since the *Q* values will be taken from the diffraction data files. It is permitted to use only a subset of the values since there is no point in calculating the EXAFS coefficients at *Q* points where no data is available; this only makes the coefficient file larger. A subset of this subset can actually be used for calculating χ^2, so the data fitted can be changed without recalculating the coefficient file. If only EXAFS data are supplied then the *Q* values are taken from this data file.

12. Other programs available

There are three other RMC codes available. RMCX fits to single crystal diffuse scattering data, RMCM uses semi-rigid molecules and RMCMAG models magnetic diffuse scattering. RMCPOW, for crystal structure refinement based on powder diffraction data (Bragg scattering), is under development.

There are many programs available in the USEFUL suite for display and analysis of the results produced by RMCA, and for the creation and modification of configurations. Programs in the EXAFS suite can be used for preparing EXAFS data for RMC modelling. Some of the NDP suite of programs are convenient for preparing experimental data sets in the correct format for input to RMCA. The MCGR program can be used to create G(*r*) from a structure factor for initial RMC modelling. All of these programs are documented separately.

13. Installing the RMC programs

13.1 Source code

The program is written in Fortran-90 and compiles under Compaq Visual Fortran as well as some other versions of the language. Most of the code is contained in one file. The configuration file reading and writing routines and some machine- dependent timing routines are found in two other files. The program is written to be general purpose and allocate array space when it runs. NB: In earlier versions the total array space available was specified by a parameter statement near the beginning of the code and needed to be changed according to the computer system being used. This restriction has now been removed with the exception that the number of background parameters is still fixed.

The configuration of atoms is stored in a file with a standard format suitable for systems containing any number of components and for molecular systems. As this file will need to be read by analysis programs it is convenient to keep the subroutines for reading and writing it separate from the main program. So that the identical code for the main program can be used on any of the machines on which we wish to run it all the machine dependent parts are contained in separate subroutines. A random number generator must be supplied. In VAX Fortran the function ran(seed) generates a random number in the interval 0 to 1. Other versions of Fortran may have equivalents. If not code for such a function is included.

13.2 Compiling and running the program

RMCA is written in FORTRAN 90 and source code exist for Windows/DOS (PC), LINUX (PC) and VMS (VAX/Alpha). The main program is self-contained except for some platform dependent routines for timing, initialisation and input/output. Plotting is done using PGPLOT routines (PGPLOT Copyright © 1983-2001 by the California Institute of Technology. http://astro.caltech.edu/~tjp/). If RMCA is downloaded as part of the WinNFLP suite of programs, then all necessary set-up would have been done following the installation procedure in the WinNFLP manual. The WinNFLP suite can be downloaded from ftp::/ftp.studsvik.uu.se/Pub/WinNFLP or http::/www.studsvik.uu.se. Otherwise the source code distribution should include the following files:

    readme.txt
    rmca.pdf
    mcgr.pdf
    rmca.for
    mcgr.for
    fileio_ WIN/UNIX/VMS.for
    rmc_ WIN/UNIX/VMS.for
    RMCLIB.for
    grfont.dat
    Makefile_LINUX *or*
    MAKEFILE_VMS.COM
    RMC_logicals.ini

Compile the relevant Fortran files and link with appropriate PGPLOT library. Sample makefiles are included with the source code and in appendix V for various compilers.
The executable distribution includes

    readme.txt
    rmca.pdf
    mcgr.pdf
    mcgr[.exe]
    rmca[.exe]
    grfont.dat
    RMC_logicals.ini

which should be extracted into a single directory,e.g. c:\nflp\rmc

For the plotting to work with VMS or LINUX, Xwindows have to run on the machine. Also, for proper plotting, the PGPLOT font and device environment variables should be defined. In Windows the RMCA program will attempt to read the RMC_logicals.ini file to define

```
PGPLOT_FONT = c:\nflp\rmc\grfont.dat
PGPLOT_DEV = /W9
```

assuming you extracted files into c:\nflp\rmc. However, if the definitions are not found at run-time then the program will attempt to find the necessary files in the program start-up directory. For LINUX version the corresponding definitions could be done in your *rlogin* file using the setenv command, also set the path to include your RMCA directory. For VMS version define a symbol to run RMCA.

```
$ rmca :== "$user$disk:[user.directory]rmca.exe"
```

This and the environment settings could be done in your login.com file. Both LINUX and VMS works best with device /xw. If PGPLOT_DEV is not defined or set to ? a list of selectable modes will be displayed at run-time

13.3 Example files:

There are some example files available which demonstrate the use of the program and which can be used for testing. The example is molten copper. The configuration is small, only 250 atoms, partly so that the program runs quickly and partly because larger files can be delayed during transfer by E-mail. The example consists of the following files:

```
cu.dat          - input data
cusq.dat        - experimental structure factor
cu.cfg          - configuration of 250 atoms
cuconv.cfg      - converged configuration
```

The file `cu.cfg` contains a suitable starting configuration which can be used to run the program. Convergence is rapid for such a small configuration. An example configuration after convergence is in `cuconv.cfg.` Note that the agreement with the experimental data is not exceptionally good at convergence because of the small size of the configuration; this is because the oscillations in g(r) have not died out by the largest distance at which g(r) can be calculated, i.e. half the box size. However this suffices as a quick example. You will find that if you use a larger configuration (say 1000-2000 atoms) it is possible to fit the experimental data very well indeed.

13.4 Using RMCA

RMCA *is not* a 'black box' program. It *will not* tell you the 'correct' structure of any material and it *will not* produce wonderful results from poor data. What RMCA will do is allow you to explore, in a controlled manner, what information on the three dimensional structure can actually be obtained from the experimental data provided. To do this it must be used in a sensible and thoughtful manner. The information obtained must also be used in an unbiased fashion. If the RMC model does not agree with a preferred theory then this should not simply be ignored. If RMC modelling produces significantly different structures that agree with the same data then this is not a fault of the method, but rather it illustrates the inadequacy of the data. The different RMC models should then be used to predict whether different, or more accurate, experiments might distinguish between the model. In this way progress can be made.

Appendix I. Example log file.

```
====================
RMCA version 3.04
====================


Using files :ZNCL2X

Reading intermediate (histogram) file ZNCL2X.his

Fraction of array space used is 10.42%

Title of run: ZnCl2 fitting to 3 neutron and 1 x-ray data sets

Configuration contains 4000 atoms of 2 types
Numbers of each type are:
                1333      2667

Configuration cell vectors are:
           25.5436    0.0000    0.0000
            0.0000   25.5436    0.0000
            0.0000    0.0000   25.5436

Number density is 0.0300 atoms per cubic Angstrom

cut offs in partial g(r)'s are at
            1-1        3.000 A
            1-2        1.800 A
            2-2        3.000 A

Maximum change in any coordinate is 0.3000A for particles of type 1
Maximum change in any coordinate is 0.3000A for particles of type 2

Using 255 r points spaced at 0.100A up to 25.500A

No. of configs to save   :   0

Writing summary every 100 generated moves
Programme will run for 595.0 minutes saving every 595.0 minutes

Fitting to 3 S(Q)´s with constant coefficients

Data set 1 uses points        4 to    96 after subtraction of 0.000
Standard deviation is 0.0030
Data will be rezeroed
Coefficients of partials are:
          0.03590 0.29420 0.60280

Data set 2 uses points        4 to    94 after subtraction of 0.000
Standard deviation is 0.0020
Data will be rezeroed
Coefficients of partials are:
          0.03590 0.22550 0.35400

Data set 3 uses points        5 to    95 after subtraction of 0.000
Standard deviation is 0.0010
Data will be rezeroed
Coefficients of partials are:
          0.03590 0.08370 0.04870

Fitting to 1 S(Q)'s with Q dependent coefficients

Data set 1 uses points        1 to 157 after subtraction of 0.000
Standard deviation is 0.0030
Data will be rezeroed
Coefficients of partials at minimum Q value are:
          0.19256 0.49251 0.31493

There are 1 coordination constraints:
    For atoms of type 2 between O.OOA and 3.OOA of atoms of type 1
        constraint is 100.0% 4-fold coordination with sigma-0.10OOE-03

There are 0 coordination number constraints:

125786 moves acc. 125913 gen. and 870036 tested; Chi-squared/d.o.f.=O.5989E+05
      Expt 1: Renorm.=1.0000;  Constant= 0.0219; Chi**2/nq= 248.8
      Expt 2: Renorm.=1.0000;  Constant= 0.0327; Chi**2/nq= 429.4
      Expt 3: Renorm.=1.0000;  Constant=-0.0026; Chi**2/nq= 574.5
      Expt 4: Renorm.=1.0000;  Constant=-0.0022; Chi**2/nq= 142.4
      Coordination constraint  1: Fraction= 49.21%; target=100.00%
```

```
125796 moves acc. 126003 gen. and 870119 tested; Chi-squared/d.o.f.=O.5971E+05
     Expt 1: Renorm.=1.0000;  Constant= 0.0219; Chi**2/nq= 248.6
     Expt 2: Renorm.=1.0000;  Constant= 0.0327; Chi**2/nq= 429.9
     Expt 3: Renorm.=1.0000;  Constant=-0.0026; Chi**2/nq= 574.7
     Expt 4: Renorm.=1.0000;  Constant=-0.0022; Chi**2/nq= 142.5
     Coordination constraint  1: Fraction= 49.29%; target=100.00%

125806 moves acc. 126116 gen. and 870220 tested; Chi-squared/d.o.f.=O.5971E+05
     Expt 1: Renorm.=1.0000;  Constant= 0.0219; Chi**2/nq= 248.6
    Expt 2: Renorm.=l.0000;  Constant= 0.0327; Chi**2/nq= 429.9
     Expt 3: Renorm.=1.0000;  Constant=-0.0026; Chi**2/nq= 574.6
     Expt 4: Renorm.=1.0000;  Constant=-0.0022; Chi**2/nq= 142.5
     Coordination constraint  1: Fraction= 49.29%; target=100.00%

125813 moves acc. 126206 gen. and 870300 tested; Chi-squared/d.o.f.=O.5954E+05
     Expt 1: Renorm.=1.0000;  Constant= 0.0219; Chi**2/nq= 248.7
     Expt 2: Renorm.=1.0000;  Constant= 0.0327; Chi**2/nq= 429.9
     Expt 3: Renorm.=1.0000;  Constant=-0.0026; Chi**2/nq= 574.6
     Expt 4: Renorm.=1.0000;  Constant=-0.0022; Chi**2/nq= 142.7
     Coordination constraint  1: Fraction= 49.36%; target=100.00%
```

Appendix II. Format of configuration file.

```
(Version 3 format configuration file)
ZnC12

125550    869706    125758 moves generated, tried, accepted
     0                     configurations saved

  4000 molecules (of all types)
     2 types of molecule
     1 is the largest numiber of atoms in a molecule
     0 Euler angles are provided

F (Box is not truncated octahedral)
  Defining vectors are:
   25.543648  0.000000   0.000000
    0.000000  25.543648   0.000000
    0.000000   0.000000  25.543648

1333 molecules of type 1
   1 atomic sites
      0.000000 0.000000 0.000000

2667 molecules of type 2
   1 atomic sites
      0.000000 0.000000 0.000000

  0.6469569  0.3505151  0.8830154
 -0.2888210 -0.6794858  0.2591100
 -0.5132699  0.9820910  0.6871831
 -0.3842981 -0.6646385 -0.5323563

 -0.3746717 -0.5287983  0.1327484
  0.9529066 -0.4303598 -0.1543689
  0.6112757  0.4258425  0.4415019
  0.6438084 -0.2311311 -0.1830015
```

Appendix III. Example data file

```
ZnC12 fitting to 3 neutron and 1 X-ray data sets
0.030                     ! number density
3.0  1.8
     3.0                  ! cut off
0.3 0.3                   ! maximum move
0.1                       ! r spacing
.false.                   ! moveout
0                         ! number of configurations to collect
100                       ! step for printing
595 595                   ! Time limit, step for saving
0 3 1 0                   ! nos of data sets
zncl235.fq
4 96                      ! points to use
0.0                       ! constant to subtract
0.0359 0.2942
       0.6028             ! coefficients
0.003                     ! sigma
.false.                   ! renormalise
.true.                    ! offset
zncl2m.fq
4 94
0.0
0.0359 0.2255
       0.3540
0.002
.false.
.true.
zncl237.fq
5 95
0.0
0.0359 0.0837
       0.0487
0.001
.false.
.true.
zncl2x.fq
1 157
0.0
0.003
.false.
.true.
1                              ! no of coordination constraints
1 2 0.0 3.0 4 1.0 0.0001   ! coordination constraint
0                              ! no of average coord constraints
.false.                   ! use a potential
```

Appendix IV. Format of experimental data files.

```
(a) S(Q) file

      157
ZnCl(35)2 neutron data

0.3000000       0.0000000E+00
0.4000000       0.0000000E+00
0.5000000       0.0000000E+00
0.6000000      -0.8116438
0.7000000      -0.7634096
0.8000000      -0.7073478
0.9000000      -0.5942892
 .
 .
 .
9.500000       -6.0869232E-02
9.600000       -3.5594340E-02
9.700001       -3.7083361E-02
9.800000       -6.2068522E-02
9.900001        0.0000000E+00
10.00000        0.0000000E+00
10.10000        0.0000000E+00
 .
 .
 .
15.60000        0.0000000E+00
15.70000        0.0000000E+00
15.80000        0.0000000E+00
15.90000        0.0000000E+00


(b) F(Q) file

       157
ZnC12 X-ray data
 0.3000000      -0.7132038    0.1925575     0.4925125      0.3149301
 0.4000000      -0.6189207    0.1936396     0.4928108      0.3135497
 0.5000000      -0.6391176    0.1950143     0.4931799      0.3118058
 0.6000000      -0.6021546    0.1966695     0.4936098      0.3097208
 0.7000000      -0.5377059    0.1985902     0.4940888      0.3073211
 0.8000000      -0.4991220    0.2007589     0.4946047      0.3046363
 0.9000000      -0.4558854    0.2031563     0.4951446      0.3016990
 .
 .
 .
15.60000      -3.1411961E-02 0.2695458      0.4992644      0.2311898
15.70000      -3.8649879E-02 0.2710147      0.4991521      0.2298333
15.80000      -3.9273359E-02 0.2724817      0.4990322      0.2284861
15.90000      -3.6237735E-02 0.2739459      0.4989051      0.2271491
```

Appendix V. Sample makefiles

Windows

LINUX (ABSOFT F90 compiler)

```
SYSTEM= $(shell uname -s)

ifeq ($(SYSTEM),Linux)
L1:= -lpgplot -L/usr/pgplot
L2:= -lX11 -L/usr/X11/lib -lm
L3:= -B108 -YEXT_NAMES="LCS" -lm -YCOM_NAMES="LCS"
L4:= -lg2c
L5:= -O -B108 -YEXT_NAMES="LCS" -lm -YCOM_NAMES="LCS"
endif

all: RMCA.o MCGR.o fileio_UNIX.o RMC_UNIX.o RMCLIB.o mcgr rmca

rmca: fileio_UNIX.o RMC_UNIX.o RMCLIB.o RMCA.o
      f90 -o rmca -l U77 fileio_UNIX.o RMC_UNIX.o RMCLIB.o RMCA.o ${L1} ${L2} ${L3}
${L4}

RMCA.o: RMCA.for
      f90 -c MCGR.for ${L3}

mcgr: fileio_UNIX.o RMC_UNIX.o RMCLIB.o MCGR.o
      f90 -o mcgr -l U77 fileio_UNIX.o RMC_UNIX.o RMCLIB.o MCGR.o ${L1} ${L2} ${L3}
${L4}

MCGR.o: MCGR.for
      f90 -c MCGR.for ${L3}

fileio_UNIX.o: fileio_UNIX.for
      f90 -c fileio_UNIX.for ${L3}

RMC_UNIX.o: RMC_UNIX.for
      f90 -c RMC_UNIX.for ${L3}

RMCLIB.o: RMCLIB.for
      f90 -c RMCLIB.for ${L3}

clean:
      rm -f *.o *.mod
```

VMS

```
$
$ L1="PGPLOT_DIR:GRPSHR.OLB/LIB"
$
$! Determine target(s)
$ target="''p1'"
$ if target.eqs.""
$ then
$  target="all"
$ else
$  goto 'target'
$ endif
$
$ all:
$
$ fileio_VMS:
$  for fileio_VMS.for
$  library/create fileio_VMS fileio_VMS
$  if target.nes."all" then exit
$
$ RMC_VMS:
$  for RMC_VMS.for
```

```
$  library/create RMC_VMS RMC_VMS
$  if target.nes."all" then exit
$
$ RMCLIB:
$  for RMCLIB.for
$  library/create RMCLIB RMCLIB
$  if target.nes."all" then exit
$
$ rmca:
$  for rmca
$  link/nouser/executable=rmca -
    rmca,RMC_VMS/lib,RMCLIB/lib,'L1'
$  if target.nes."all" then exit
$
$ mcgr:
$  for mcgr
$  link/nouser/executable=mcgr -
    mcgr,RMC_VMS/lib,RMCLIB/lib,'L1'
$  if target.nes."all" then exit
$
$! Clean only if not "all"
$ exit
$
$ clean:
$  delete/noconfirm *.obj;*,*.olb;*,*.f90$mod;*
$  exit
```

RMC bibliography.

1 N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller
   J. Phys. Chem. **21** 1087

2 Atomic arrangement in vitreous selenium. R. Kaplow, T.A. Rowe and B.L. Averbach.
   1968 *Phys. Rev.* **168** 1068

3 Monte Carlo models of atomic arrangements in arsenic-selenium glass. A.L. Renninger, M.D. Rechtin and
   B.L. Averbach. 1974 *J. Non-Cryst. Solids* **16** 1

4 Reverse Monte Carlo simulation: a new technique for the determination of disordered structures. R L
   McGreevy and L Pusztai. 1988 *Mol. Simulation.* **1** 359-367

5 The structure of molten LiCl. R L McGreevy and M A Howe.
   1989 *J. Phys: Condensed Matter* **1** 9957-62

6 The structure of liquid copper selenide. M.A. Howe.
   1989 *Physica* **B160** 170

7 Structure modelling of Molecular liquids for the systems Benzene, Hexafluorobenzene and their
   equimolar mixture. M. Ostheimer and H. Bertagnolli. 1989.
   Z. Phys. Chem. Neue Folge **162** 171

8 Reverse Monte Carlo simulation techniques for combining X-ray and neutron diffraction data. R L
   McGreevy. 1989 *IOP conference series* **101** 41-50

9 Orientational correlations in the liquid halogens. M.A. Howe.
   1990 *Mol. Phys.* **69** 161

10 The hydration of ions in aqueous solution: reverse Monte Carlo analysis of neutron diffraction data. M.A.
   Howe. 1990 *J. Phys: Condensed Matter* **2** 241

11 Comments on reverse Monte Carlo simulation. R.A. Evans.
   1990 *Mol. Sim.* **4** 409

12 Structural disorder in AgBr at the approach to melting. D A Keen, R L McGreevy and W Hayes. *1990 J.
   Phys: Condensed Matter* **2** 2773-2786

13 Structural disorder in AgBr: RMC analysis of powder neutron diffraction data. D A Keen, W Hayes, R L
   McGreevy and K N Clausen. 1990 *Phil. Mag. Lett.* **61**
   349-357

14 Structural modelling of glasses using reverse Monte Carlo simulation. D A Keen and R L McGreevy.
   1990 *Nature* **344** 423-42

15 The structures of molten salts. R L McGreevy and L Pusztai.
   1990 *Proc. Roy. Soc. A* **430** 241-261

16 The structure of glassy Zinc Chloride: A reverse Monte Carlo study. L Pusztai and R L McGreevy. 1990
   J. Non-Cryst. Solids **117-118** 627-628

17 Reverse Monte Carlo simulation for the analysis of EXAFS data. S J Gurman and R L McGreevy *1990 J.
   Phys: Condensed Matter* **2** 9463-9474

18 Reverse Monte Carlo (RMC) simulation: modelling structural disorder in crystals, glasses and liquids
   from diffraction data. R L McGreevy, M A Howe, D A Keen and K N Clausen. *1990 IOP
   conference series* **107** 165-184

19 The structure of molten K-Pb, Rb-Pb and Cs-Pb alloys. M A Howe and R L McGreevy. 1991 *J. Phys:
   Condensed Matter* **3** 577-591

20 The metal-non metal transition in expanded caesium. V M Nield, M A Howe and R L McGreevy. *1991 J.
   Phys: Condensed Matter* **3** 7519-7525

21 Determination of three body correlations in liquids by RMC modelling of diffraction data. I. Theoretical
   tests. M A Howe and R L McGreevy. 1991 *Phys. Chem. Liq.* **24** 1-12

22 Determination of disordered magnetic structures by RMC modelling of neutron diffraction data. D A
   Keen and R L McGreevy. 1991 *J. Phys: Condensed Matter* **3** 7383- 7394

23 The structure of liquid "He across the A-point. L Pusztai and R L McGreevy. 1991
   Phys. Chem. Liq. **24** 119-125

24 The structure of a-C:H by neutron and X-ray scattering. R.J. Newport, P.J.R. Honeybone, S.P. Cottrell, J. Franks, P. Revell, R.J. Cernik and W.S. Howells. 1991 *Surf. Coat. and Tech.* **47** 608

25 Fractal aspects of superionic glasses from Reverse Monte Carlo simulations. L Börjesson, R L McGreevy and W S Howells. 1992 *Phil. Mag. B* **65** 261-271

26 Neutron scattering studies of disorder and melting in Ag and Cu halides. D A Keen, V M Nield, R L McGreevy and W Hayes. 1992 *Physica B* **180-181** 798-801

27 Determining the structures of disordered materials by diffuse neutron scattering and RMC modelling. R L McGreevy, M A Howe, V M Nield, J D Wicks and D A Keen. 1992 *Physica, B* **180-181** 801-804

28 RMC: modelling disordered structures. R L McGreevy and M A Howe. 1992 *Annual Review of Materials Science* **22** 217-242

29 Structural disorder in AgBr at the approach to melting. IL V M Nield, D A Keen, W Hayes and R L McGreevy. 1992 J. *Phys: Condensed Matter* **4** 6703-6714

30 The structure of the C_{70} molecule. D.R. McKenzie, C.A. Davis, D.J.H. Cockayne, D.A. Muller and A.M. Vassallo. 1992 *Nature* **355** 622

31 The structure of superionic glasses from neutron diffraction and reverse Monte Carlo modelling. L. Börjesson, R.L. McGreevy and J.D. Wicks. 1992 *J. de Phys. IV* **C2** 107

32 Determination of three body correlations in liquids by RMC modelling of diffraction data. IL Elemental liquids. M A Howe, R L McGreevy, L Pusztai and I Borzsak. 1993 *Phys. Chem. Liq.* **25** 204

33 RMC techniques for structural modelling. R L McGreevy. 1993 *Adv. Solid State Chem. in press*

34 Neutron diffraction and RMC modelling: complementary techniques for the determination of disordered structures. R L McGreevy. 1993 *Int. J. Mod. Phys. B* **7** 2965

35 RMC: fact and fiction. R.L. McGreevy. 1993 *J. Non-Czyst. Solids* **156-158** 949

36 The determination of the structure of $(AgI)_x(AgPO3)_{1-x}$ glass by a combination of X-ray and neutron diffraction data. J.D. Wicks, R.L. McGreevy and L. Börjesson. 1993 *Int. J. Mod. Phys. B in press*

37 RMC techniques for the determination of partial structure factors. R.L. McGreevy. 1993 *Int. J. Mod. Phys. B in press*

38 A reverse Monte Carlo study of amorphous $Ti_{67}Ni_{33}$. E.W. Iparraguirre, J. Sietsma and B.J. Thijsse. 1993 *J. Non-Cryst. Solids* **156-158** 969

39 The performance of the Reverse Monte Carlo method applied to the metallic glasses $Ni_{81}B_{19}$ and $Ti_{67}Ni_{33}$. E.W. Iparraguirre, J. Sietsma and B.J. Thijsse. 1993 *Int. J. Mod. Phys. B in press*

40 A reverse Monte Carlo study of amorphous $Ni_{81}B_{19}$. E.W. Iparraguirre, J. Sietsma and B.J. Thijsse. 1993 *Comp. Phys. Comm. in press*

41 C. Back, J.M. Dubois and R. Bellissent. RMC simulation of a metallic glass. 1993 *Int. J. Mod. Phys. B in press*

42 RMC modelling of the structure of liquid lithium-silicon alloys. P.H.K. de Jong, P. Verkerk and L.A. de Graaf. 1993 *Int. J. Mod. Phys. B in press*

43 The structure of liquid lithium-silicon alloys. P.H.K. de Jong, P.Verkerk. W. van der Lugt, W.S. Howells and L.A. P. de Jong 1993 *J. Non-Cryst. Solids* **156-158** 978

44 RMC experimentation on a metal-meta-I amorphous alloy. M. Bionducci. F. Buffa, G. Licheri and G. Navarra. 1993 *Int. J. Mod. Phys. B in press*

45 Amorphous CuZr alloy investigated by anomalous X-ray scattering. M. Bionducci, F. Buffa, G. Licheri and G. Navarra. *1993 Int. J. Mod. Phys. B in press*

46 Structure study of Ni62Nb38 metallic glass using reverse Monte Carlo simulation. L. Pusztai and E. Svib. 1993 *J. Non-Czyst. Solids* **156-158** 973

47 A complete atomistic model of molten poly(ethylene) from neutron scattering data: a new methodology for polymer structures. B. Rosi-Schwaxtz and G.R. Mitchell. 1993 *submitted to Makromol. Chemie - Theory and Simulation*

48 I.D. Brown, in Structure and Bonding in Crystals, vol.II, eds. M. O'Keeffe and A.Navrotsky, Academic Press, New York, USA, pp 1-30.

The structure of liquid sulphur: an RMC case study

L Karlsson and R L McGreevy

October 1993

Introduction

The aim of this paper is to describe, in considerable detail, how the reverse Monte Carlo method and the RMCA (version 3) program (and related programs) can be applied to the results of a particular diffraction experiment. Hopefully the reader will then understand how to apply the program(s) to their own data in a logical and efficient manner.

Liquid sulphur has been chosen for this case study because in some senses it is a simple system, i.e. it is a monatomic liquid, and in other senses it is extremely complex, that is the structure may be molecular or polymeric. Because of these 'contradictory' aspects it is in fact probably the most scientifically difficult and challenging study yet undertaken using RMC. The data is of very high quality; this avoids some of the problems that can be encountered if the available data is inaccurate and which are hard to quantify without considerable experience (which has nothing to do with RMC modelling). The experimental structure factor is shown in figure 1.

The modelling procedure will be described in the next section, step by step. The data files for input to RMCA and other programs are given in order in the appendices and the results are shown in the figures.

MCGR modelling

The first step is to obtain the radial distribution function, G(r). It saves considerable time to fit G(r) first, and some information from G(r) such as closest approach distances is also useful to know in advance. `MCGR` is a program for obtaining the radial distribution function by an inverse method; see the `MCGR` documentation for full details.

Step 1.

The original structure factor, shown in figure 1, contains 987 points from 0.52 to 39.96 Å^{-1} at 0.04 Å^{-1} intervals. G(r) obtained from this by direct Fourier transform (using NDP routine GOFR) is shown in figure 2. The first peak is easily identified as being at 2Å, and a first guess at the closest approach distance is 1.7Å.

Step 2.

The full Q resolution and range is not necessary for initial MCGR fitting so we create a new structure factor (figure 3) with 145 Q points from 0.6 to 15.0 Å^{-1} at 0.1 Å^{-1} intervals (using NDP routine REBIN). MCGR is run with parameters given in Appendix I. Since there are still weak oscillations in F(Q) at 30 Å^{-1} an appropriate choice of r spacing is close to $2\pi/(5 \times 30)$, so we have used 0.05 Å. From G(r) obtained by direct transform it appears that a maximum r of 20 Å is sufficient. The closest approach cut-off of 1.7Å, as guessed in step 1,

is used. The resulting fit to the structure factor is shown in figure 4 and the corresponding G(r) in figure 5. The final χ^2 is 1.5. The G(r) suggests a new closest approach value of 1.9Å. The *r* spacing and maximum *r* used seem to be appropriate.

Step 3.

A new structure factor with 300 *Q* points from 0.6 to 30.5 Å^{-1} is created. MCGR is run again with a cut-off of 1.9Å (parameters in Appendix II). The fit is shown in figure 6 and the resulting G(r) in figure 7. The final χ^2 is 1.5. The first peak in G(r) is asymmetric so the cut-off used is probably too high.

Step 4.

Run MCGR again, this time with cut-off 1.85Å (Appendix III). The fit is shown in figure 8 and G(r) in figure 9. The final χ^2 is 1.2. The first peak in G(r) is now more symmetric.

Step 5.

Now run MCGR with the same parameters, but saving 20 sets of G(r) to average over. The target χ^2 for saving is 1.2 (Appendix IV). The resulting G(r) (average and standard deviation) is shown in figure 10. This average will initially be modelled by RMC.

RMC modelling (no constraints)

Step 6.

First the initial configuration must be created for the RMC model. Since the material studied is a liquid, and we are using no prior information (i.e. no constraints) the initial configuration chosen is random. First run the program RMC:RANDOM (Appendix V) to create a random set of 4000 atomic positions. 4000 atoms is typically found to be a reasonable number unless there are particular requirements for larger configurations.

Step 7.

To save time the random model should be made consistent with the atomic closest approach that is going to be used. This can be acheived most quickly, though not always completely if the packing fraction is high, using the program RMC:MOVEOUT. The best way to run this program is by moving the atoms apart in small steps. In this case we can start at 1.05Å and increase this to 1.85Å in 0.1Å steps, allowing 0.2Å moves (see Appendix VI). G(r) for the resulting configuration (not the same as an equilibrated hard sphere simulation, though close) is shown in figure 11, in comparison to the target G(r) created by MCGR in step 5.

Step 8.

Now we run RMC, starting from the random configuration generated in step 7. The parameters are given in Appendix VII The initial χ^2 is 1430. After 24 hours (VAX4000) it is 11.8. The fit is shown in figure 12.

Step 9.

σ is decreased to 0.005 (Appendix VIII). The initial χ^2 is now 47.1. After 50 hours (VAX4000) this has decreased to 20.5. The fit is shown in figure 13; it can be seen that this is a good fit

Step 10.

Now we can start to fit the structure factor. We will use all of the available Q range, but the Q resolution of 0.04Å^{-1} is too good considering that the maximum r value in the RMC model is 24.6Å. We make a new F(Q) with 393 Q points from 0.6 to 39.8 Å^{-1} at 0.1Å^{-1} intervals. σ is chosen to be 0.0025. The initial χ^2 is 25.94. After 40 hours (VAX4000) this has decreased to 6.734. The fit is shown in figure 14 and the parameters are given in Appendix IX. The resulting G(r) is compared in figure 15 with that from MCGR (step 5).

Step 11.

Since a good fit has already been achieved we can try to get a slight further improvement by allowing the data to renormalise (Appendix X). The final χ^2 is 6.5 so there is no significant improvement; the configuration from step 10 is kept.

Step 12.

We can now make some initial analysis of the RMC model. From G(r) (figure 15) we find a well defined minimum at 2.5Å. We run RMC:TRIPLETS (Appendix XI) and RMC:NEXTTO (Appendix XII) to get the bond angle (figure 16) and neighbour distributions (Appendix XIII) in the first coordination shell. The average coordination number is 1.98 and the neighbour distribution peaks at 2. From what is known about the chemistry of sulphur there is then a strong possibility that constraints could be used to impose two-fold coordination. This leads to the next step.

RMC modelling (coordination constraints)

Step 13.

The approach is now to try to create a model with the maximum amount of twofold coordination, representing preferred covalent bonding. Initially we can remove atoms with high coordination (4 and above), and low coordination (0), by applying coordination constraints that try to reduce the number of atoms with such coordinations to zero (Appendix XIV). The fit to the data is not as good as before, the χ^2 being 21.84. The new neighbour distribution is given in Appendix XV.

Step 14.

We now try to produce a configuration where all atoms have two-fold coordination (Appendix XVI). Initially the two-fold coordination increases quickly, but then progress becomes increasingly slow. After 20 hours (VAX4000) 84.35% have two-fold coordination; after 200 hours this has increased to 93.97%. Eventually there is little point in continuing. The fit to the data is still as good (χ^2 is 21.1) but now 94% of S atoms are two-fold coordinated. If further improvement was wanted it would now be possible to alter the positions of the few remaining atoms with the wrong coordination by 'hand' (suitable programs are available). However it should be noted that this will really make very little difference to any of the properties of the structure, and so is only worthwhile if considered absolutely necessary.

Step 15.

The neighbour distribution for the coordination constrained configuration is given in Appendix XVII and the bond angle distribution is shown in figure 17.

Step 16.

One possible structure for liquid sulphur is that it consists of S_8 ring molecules. It is extremely unlikely that many S_8 rings will form 'by themselves' in an RMC model, since statistically there are clearly many other possible arrangements which would agree with the data. In order to test the possibility of S_8 rings we must therefore start with them already in the model. Here we can do this by starting from the crystal structure of α-S which contains S_8 rings. All atoms are in rings so all atoms are 2-fold coordinated. By using a constraint of 100% 2-fold coordination we can therefore keep all of these rings connected, while allowing them to move around in order to try to fit the data for liquid S, i.e. coordination constraints are being used to produce flexible molecules. Because the molecules in this case are so large it takes a very long time for them to lose the 'memory' of the initial crystal structure. We therefore run RMCA first without fitting to any data (Appendix XVIII) to disorder the inter-molecular structure, and then fit to the data (Appendix XIX). This takes a very long time.

The resulting fit is shown in figure 18. The bond angle distribution is shown in figure 19. χ^2 is 5.9, a better fit than was achieved with 2-fold coordination constraints before (Step 14). This may have a physical significance, or it may have another cause. The bond angle distributions in figures 17 and 19 differ in that one has a peak at 60°, while the other does not. This peak is due to the presence Of S3 triangles in one model. All the atoms in a triangle have the same separation and 2-fold coordination, so they can fit a sharp first peak in (r) and a 2-fold coordination constraint well. S_3 triangles are highly unlikely in liquid S from basic chemical bonding considerations, so they should be removed from the model. This requires monitoring of three body correlations in RMC which is very expensive. However a trial constraint of this form has been included in a development version of RMCA.

Conclusions

We have illustrated briefly how the RMC method, and RMCA program, can be used in a particular structural study. The last part of this study is deliberately inconclusive. By themselves the diffraction data are almost certainly insufficient to resolve the different possible structures of liquid sulphur, so RMC modelling can only give some guidance as to the range of possibilities. These possibilities must then be considered in the light of other information. If this is real information (rather than interpretation) then it may be possible to use it in the form of a constraint in the RMC model (for example the absence of S_3 triangles), which may then require further program development.

Figure 1. The experimental structure factor for liquid sulphur at 140°C.

Figure 2. G(r) obtained from the experimental structure factor by direct Fourier transform. (Step 1)

Figure 3. Rebinned structure factor used for inital MCGR fitting. (Step 2)

Figure 4. Initial MCGR fit (broken line) compared to the experimental data (solid line). (Step 2)

Figure 5. G(r) corresponding to the initial MCGR fit. (Step 2)

Figure 6. MCGR fit (broken line) to the experimental structure factor (solid line) with the G(r) cut-off at 1.9 Å. (Step 3)

Figure 7. G(r) corresponding to the MCGR fit with the cut-off at 1.9 Å. (Step 3)

Figure 8. Final MCGR fit (broken line) to the experimental structure factor (solid line) with the G(r) cut-off at 1.85 Å. (Step 4)

Figure 9. G(r) corresponding to the MCGR fit with the cut-off at 1.85 Å. (Step 4)

Figure 10. Average and standard deviation of G(r) from the final MCGR fit. (Step 5)

Figure 11. G(r) for the initial random RMC configuration (solid line) compared to the MCGR result (broken line). (Step 7)

Figure 12. Initial RMC fit (broken line) to the G(r) from MCGR (solid line). (Step 8)

Figure 13. Final RMC fit (broken line) to the G(r) from MCGR (solid line). (Step 9)

Figure 14. Final RMC fit (broken line) to the experimental structure factor (solid line), with no coordination constraints. (Step 10)

Figure 15. G(r) from the RMC model with no coordination constraints (solid line) compared to the MCGR result (broken line). (Step 10)

Figure 16. Bond angle distribution for the RMC model with no coordination constraints. (Step 12)

Figure 17. Bond angle distribution for the RMC model with 2-fold coordination constraint. (Step 15)

Figure 18. RMC fit (broken line) to the experimental structure factor (solid line) for a flexible S_8 ring molecule model. (Step 16)

Figure 19. Bond angle distribution for the RMC model with flexible S_8 ring molecules. (Step 16)
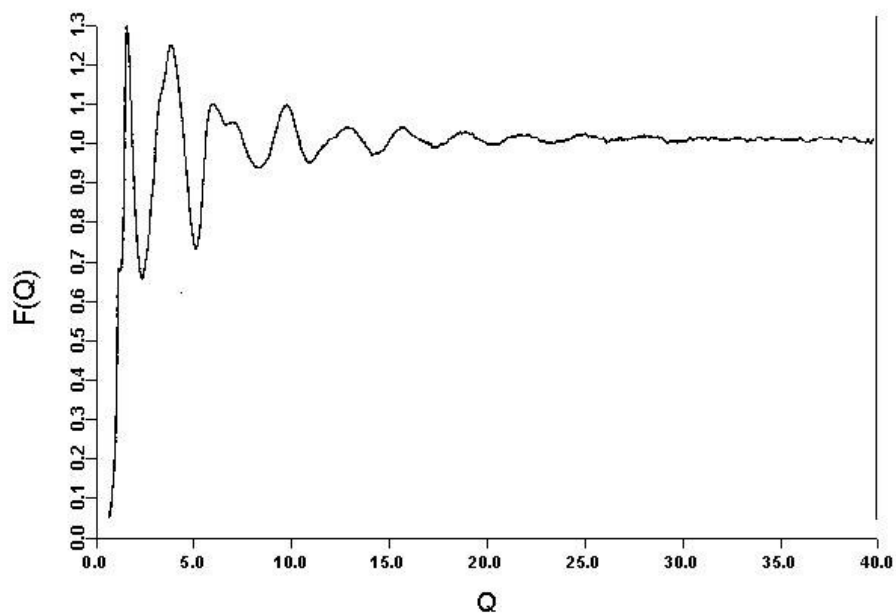
Figure 1. The experimental structure factor for liquid sulphur at 140°C.
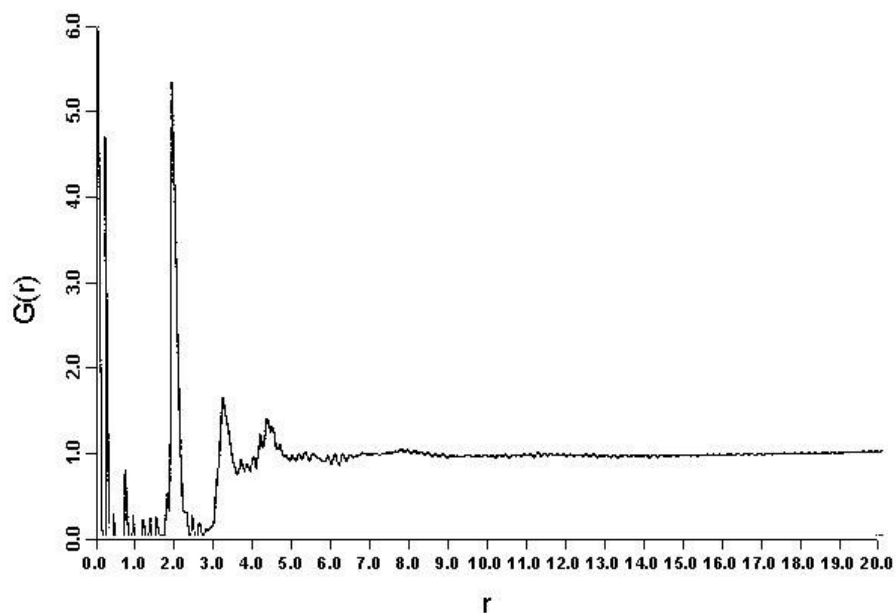

Figure 2. G(r) obtained from the experimental structure factor by direct Fourier transform. (Step 1)
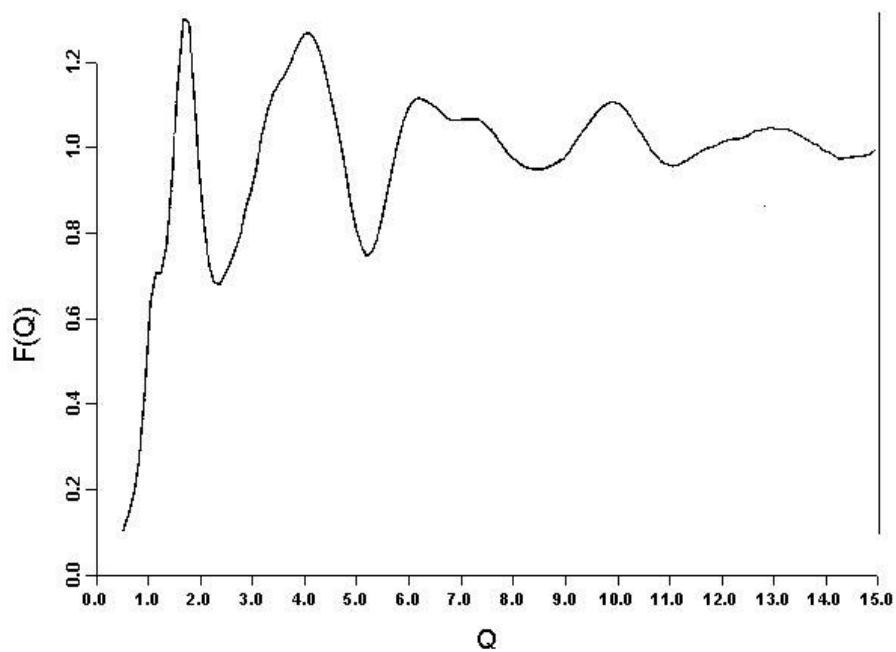
Figure 3. Rebinned structure factor used for initial MCGR fitting. (Step 2)
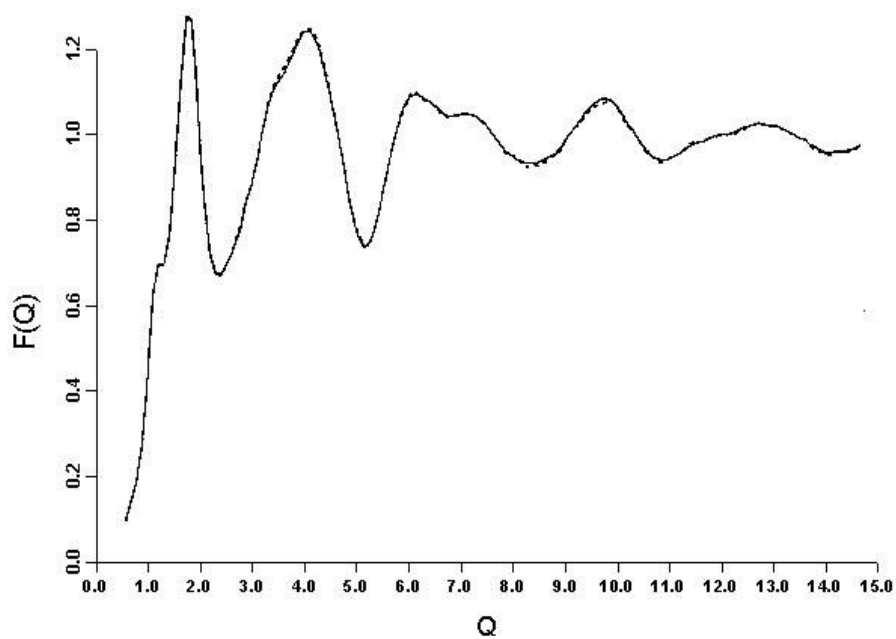

Figure 4. Initial MCGR fit (broken line) compared to the experimental data (solid line). (Step 2)
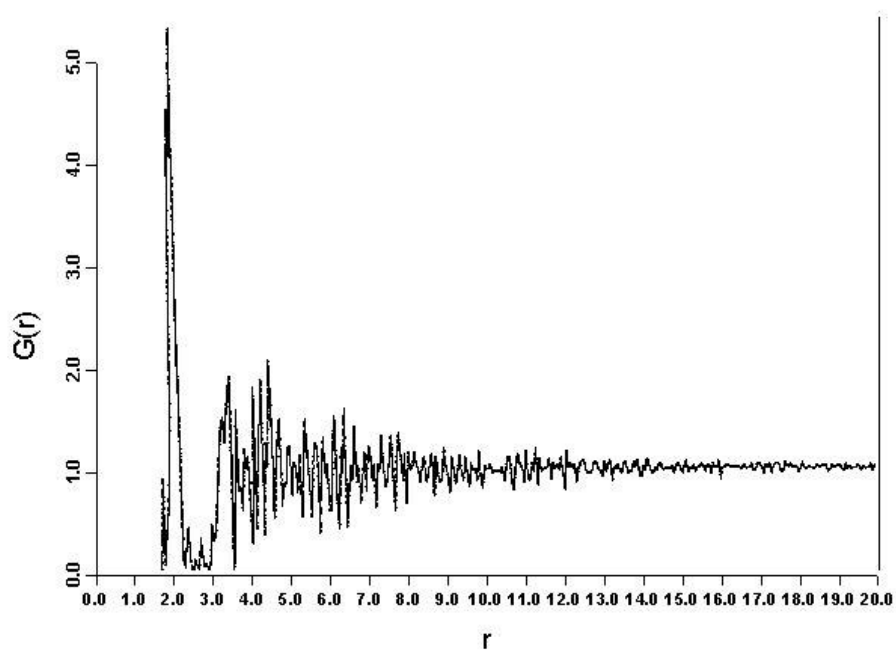
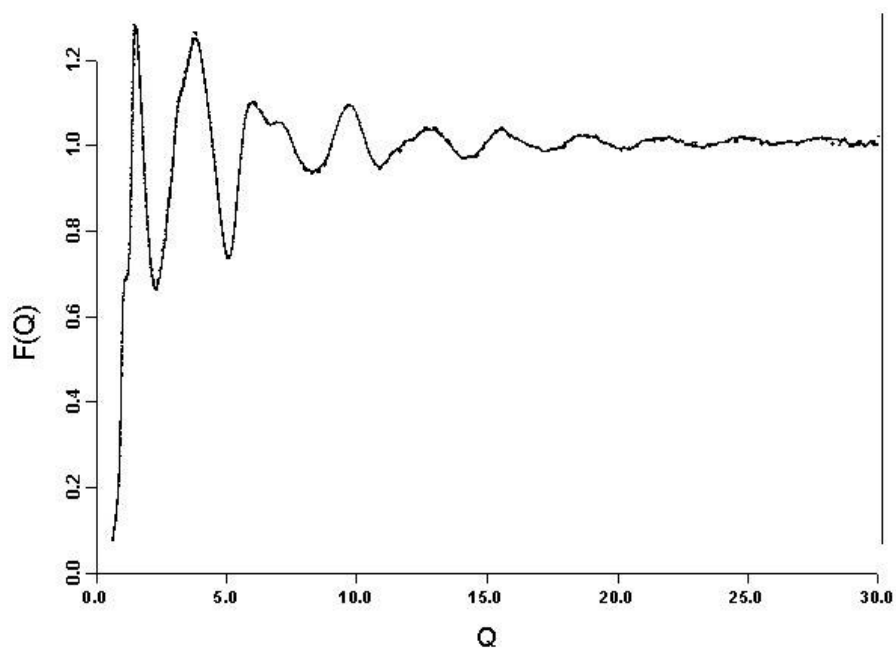Figure 5. G(*r*) corresponding to the initial MCGR fit (Step 2)

Figure 6. MCGR fit (broken line) to the experimental structure factor (solid line) with the G(*r*) cut-off at 1.9 Å. (Step 3)
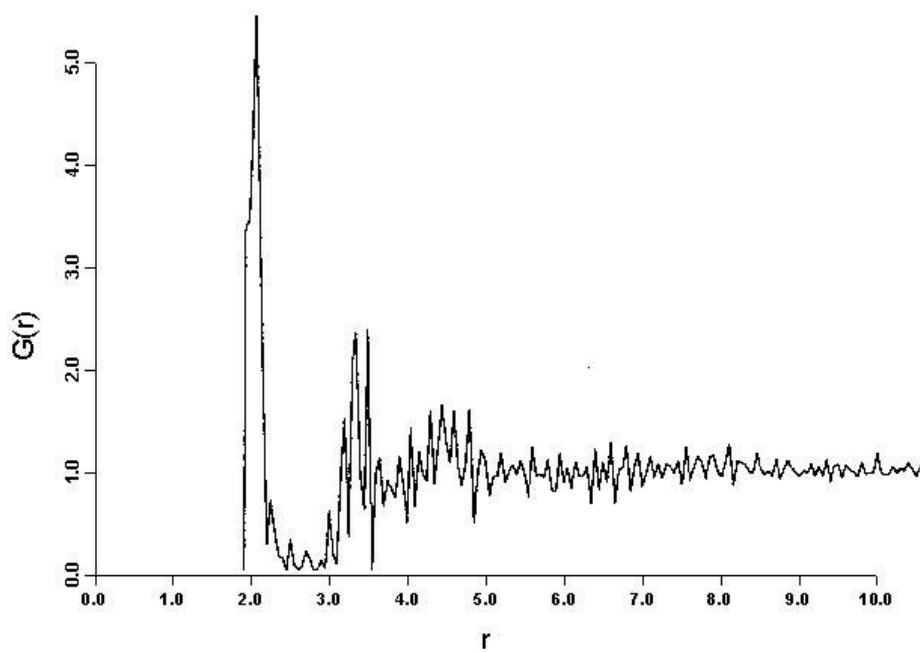
Figure 7. G(r) corresponding to the MCGR fit with the cut-off at 1.9 Å. (Step 3)
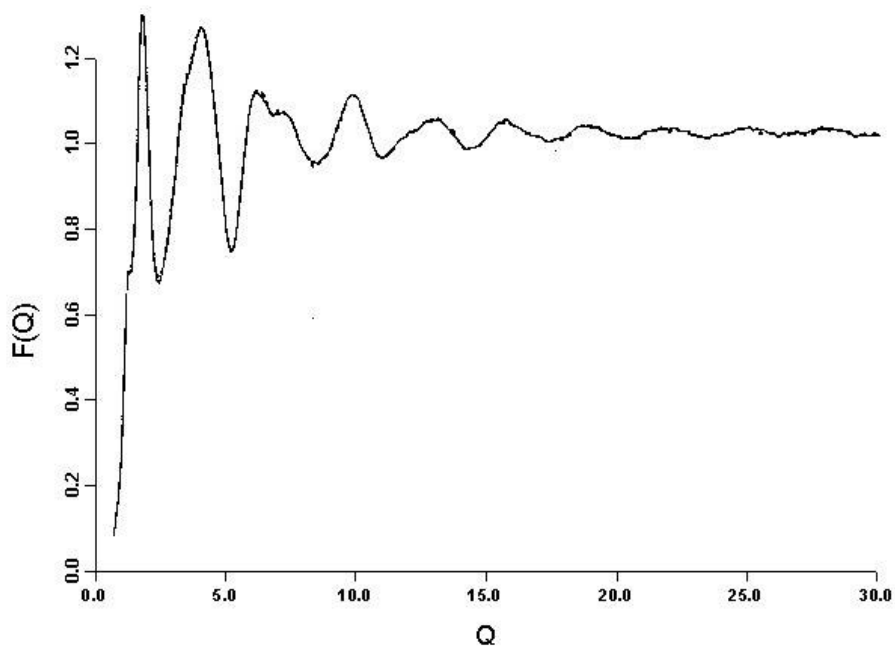

Figure 8. Final MCGR fit (broken line) to the experimental structure factor (solid line) with the G(r) cut-off at 1.85 Å. (Step 4)
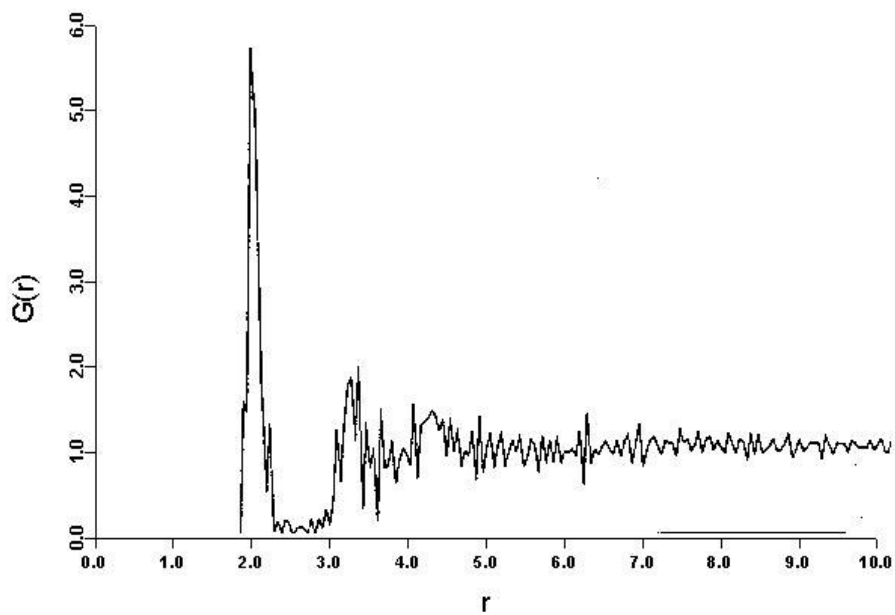
Figure 9. G(*r*) corresponding to the MCGR fit with the cut-off at 1.85Å. (Step 4)

Figure 10. Average and standard deviation of G(*r*) from the final MCGR fit (Step 5)

Figure 11. G(*r*) for the initial random RMC configuration (solid line) compared to the MCGR result (broken line). (Step 7)

Figure 12. Initial RMC fit (broken line) to the G(*r*) from MCGR (solid line). (Step 8)

Figure 13. Final RMC fit (broken line) to the G(r) from MCGR (solid line). (Step 9)

Figure 14. Final RMC fit (broken line) to the experimental structure factor (solid line), with no coordination constraints. (Step 10)

Figure 15. G(*r*) from the RMC model with no coordination constraints (solid line) compared to the MCGR result (broken line). (Step 10)

Figure 16. Bond angle distribution for the RMC model with no coordination constraints. (Step 12)

Figure 17. Bond angle distribution for the RMC model with 2-fold coordination constraint. (Step 15)

Figure 18. RMC fit (broken line) to the experimental structure factor (solid line) for a flexible S_8 ring molecule model. (Step 16)

Figure 19. Bond angle distribution for the RMC model with flexible S_8 ring molecules. (Step 16)

Appendix I. Parameters for initial MCGR fitting

```
S1mcgr
0.0335          ! number density
.false.         ! generate partials
1               ! no. of partials
1.7             ! cut off
0.1             ! delta
400,0.05        ! r points, spacing
.false.         ! save
.false.         ! converge only
.true.          ! g(r) positive
.false.         ! smoothing
.false.         ! resolution
1000 0          ! step for printing, plotting off
10 5            ! time limit, step for saving
1 0             ! no. of data sets
s1_0.fq
1 150           ! points to fit
1               ! constant to subtract
1               ! coefficients
0.005           ! sigma
.false.         ! renom
.true.          ! offset
.false.         ! magn.
```
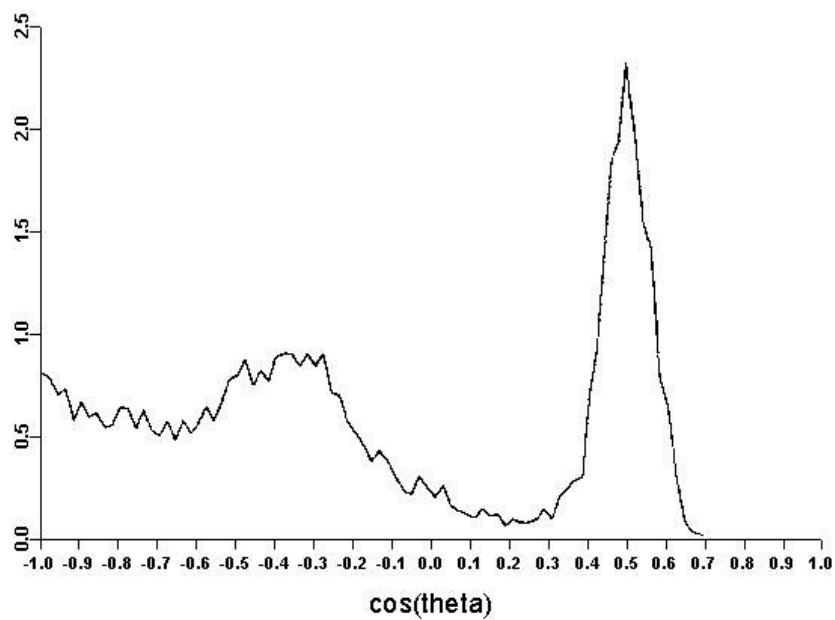

Appendix II. Parameters for MCGR, step 3

```
S1mcgr
0.0335          ! number density
.false.         ! generate partials
1               ! no. of partials
1.9             ! cut off
0.1             ! delta
400,0.05        ! r points, spacing
.false.         ! save
.false.         ! converge only
.true.          ! g(r) positive
.false.         ! smoothing
.false.         ! resolution
1000 0          ! step for printing, plotting off
10 5            ! time limit, step for saving
1 0             ! no. of data sets
s1_1.fq
1 300           ! points to fit
1               ! constant to subtract
1               ! coefficients
0.005           ! sigma
.false.         ! renom
.true.          ! offset
.false.         !magn.
```

```
Appendix III. Parameters for MCGR, step 4

S1mcgr
0.0335          ! number density
.false.         ! generate partials
1               ! no. of partials
1.85            ! cut off
0.1             ! delta
400,0.05        ! r points, spacing
.false.         ! save
.false.         ! converge only
.true.          ! g(r) positive
.false.         ! smoothing
.false.         ! resolution
1000 0          ! step for printing, plotting off
20 10           ! time limit, step for saving
1 0             ! no. of data sets
s1_1.fq
1 300           ! points to fit
1               ! constant to subtract
1               ! coefficients
0.005           ! sigma
.false.         ! renom
.true.          ! offset
.false.         ! magn.




Appendix IV. Parameters for final MCGR run, saving 20 sets of G(r)

S1mcgr
0.0335          ! number density
.false.         ! generate partials
1               ! no. of partials
1.85            ! cut off
0.1             ! delta
400,0.05        ! r points, spacing
.true.          ! save
1.2, 20
.false.         ! converge only
.true.          ! g(r) positive
.false.         ! smoothing
.false.         ! resolution
1000 0          ! step for printing, plotting off
60 10           ! time limit, step for saving
1 0             ! no. of data sets
s1_1.fq
1 300           ! points to fit
1               ! constant to subtract
1               ! coefficients
0.005           ! sigma
.false.         ! renorm
.true.          ! offset
.false.         ! magn.
```

```
Appendix V. A RMC:RANDOM run for creation of a random configuration

r rmc:random
Truncated octahedral ? (t/f) >
f
Number of Euler angles       >
0
Number of particle types     >
1
Density                      >
0.0335
Number of particles of type 1>
4000
Output file                  >
rand.cfg
FORTRAN STOP


Appendix VII. Parameters for RMC, fitting G(r)

S1
0.0335              ! number density
  1.85              ! cut offs
0.2                 ! maximum move
0.05                ! r spacing
.false.             ! whether to use moveout option
0                   ! number of configurations to collect
1000 0              ! step for printing, plotting off
60 60               ! Time limit, step for saving
1 0 0 0             ! No. of g(r), neutron, X-ray, EXAFS expts
slmcgr30.gr
1 400               ! Range of points
0                   ! Constant to subtract
1                   ! Coefficients
0.01                ! Standard deviation
.false.             ! Renormalise
0                   ! no. of coordination constraints
0                   ! no. of average coordination constraints
.false.             ! whether to use a potential
```

```
Appendix VI. A RMC:MOVEOUT run

r rmc:moveout
Starting configuration>
rand.cfg
Closest approaches   >
1.05
 592 atoms have too close neighbours
maximum move         >
1
max. no. of iterations>
1000
 591 atoms too crowded; 1 failed moves
 590 atoms too crowded; 1 failed moves
 589 atoms too crowded; 1 failed moves
  .
  .
  .
   2 atoms too crowded; 415 failed moves
   1 atoms too crowded; 416 failed moves
   0 atoms too crowded; 418 failed moves
Change cut-offs ?    >
t
Closest approaches   >
1.15
 238 atoms have too close neighbours
Maximum move         >
0.2
max. no. of iterations>
1000
  237 atoms too crowded; 1 failed moves
  236 atoms too crowded; 4 failed moves
  .
  .
  .
  .
  .
Change cut-offs ?     >
t
Closest approaches    >
1.85
1533 atoms have too close neighbours
Maximum move         >
0.2
max. no. of iterations>
6000
1532 atoms too crowded; 0 failed moves
1531 atoms too crowded; 1 failed moves
  .
  .
  .
   1 atoms too crowded; 5461 failed moves
   0 atoms too crowded; 5468 failed moves
Change cut-offs       >
f
Output file           >
moveout.cfg
FORTRAN STOP
```

```
Appendix VIII. Parameters for RMC, fitting G(r) with decreased sigma

S1
0.0335              ! number density
  1.85              ! cut offs
0.2                 ! maximum move
0.05                ! r spacing
.false.             ! whether to use moveout option
0                   ! number of configurations to collect
1000 0              ! step for printing, plotting off
60 60               ! Time limit, step for saving
1 0 0 0             ! No. of g(r), neutron, X-ray, EXAFS expts
s1mcgr30.gr
1 400               ! Range of points
0                   ! Constant to subtract
1                   ! Coefficients
0.005               ! Standard deviation
.false.             ! Renormalise
0                   ! no. of coordination constraints
0                   ! no. of average coordination constraints
.false.             ! whether to use a potential


Appendix IX. Parameters for RMC, fitting F(Q)

S1
0.0335              ! number density
  1.85              ! cut offs
0.1                 ! maximum move
0.05                ! r spacing
.false.             ! whether to use moveout option
0                   ! number of configurations to collect
1000 0              ! step for printing, plotting off
60 60               ! Time limit, step for saving
0 1 0 0             ! No. of g(r), neutron, X-ray, EXAFS expts
S1_2.fq
1 500               ! Range of points
0                   ! Constant to subtract
1                   ! Coefficients
0.0025              ! Standard deviation
.false.             ! Renormalise
.true.              ! Offset
0                   ! no. of coordination constraints
0                   ! no. of average coordination constraints
.false.             ! whether to use a potential
```

```
Appendix X. RMC parameters, allowing data to renormalise

S1
0.0335              ! number density
  1.85              ! cut offs
0.1                 ! maximum move
0.05                ! r spacing
.false.             ! whether to use moveout option
0                   ! number of configurations to collect
1000 0              ! step for printing, plotting off
60 60               ! Time limit, step for saving
0 1 0 0             ! No. of g(r), neutron, X-ray, EXAFS expts
s1_2.fq
1 500               ! Range of points
0                   ! Constant to subtract
1                   ! Coefficients
0.0025              ! Standard deviation
.true.              ! Renormalise
.true.              ! Offset
0                   ! no. of coordination constraints
0                   ! no. of average coordination constraints
.false.             ! whether to use a potential


Appendix XI. A RMC:TRIPLETS run, to get bond angle distribution

r rmc:triplets
No. of theta pts                                >
100
No. of neighbours for bond ang (0 for all) >
0
Number of configurations                        >
1
Configuration file                              >
S1_4.cfg
Maximum r values                                >
2.5
Output file                                     >
s1.trip
FORTRAN STOP


Appendix XII. A RMC:NEXTTO run, to get neighbour distribution

r rmc:nextto
Configuration file              >
S1_4.cfg
Distances from central particle >
2.5
Output file
S1.nex
FORTRAN STOP
```

Appendix XIII. Coordination number distribution, no constraints

Output from NEXTTO
TITLE
S1
Coordination number distributions

central particle: 1
Distance: 2.500A
Coordination number distributions
BARCHART
CURVES
  7 1
             0   251

             1   1099

             2   1483
             3   863
             4   248
             5   48
             6   8
Coordination numbers:
  1.983500

Appendix XIV. RMC parameters, modelling with coordination constraints

S1r
0.0335   ! number density
  1.85   ! cut offs
0.5   ! maximum move
0.05   ! r spacing
.false.   ! whether to use moveout option
0   ! number of configurations to collect
1000 0   ! step for printing, plotting off
60 60   ! Time limit, step for saving
0 1 0 0   ! No. of g(r), neutron, X-ray, EXAFS expts
s1_2.fq
1 500   ! Range of points
0   ! Constant to subtract
1   ! Coefficients
0.0025   ! Standard deviation
.false.   ! Renormalise
.true.   ! Offset
4   ! no. of coordination constraints
1 1 0 2.5 0 0 le-5
1 1 0 2.5 4 0 le-5
1 1 0 2.5 5 0 le-5
1 1 0 2.5 6 0 le-5
0   !no. of average coordination constraints
.false.   ! whether to use a potential

Appendix XV. Coordination number distribution, with coordination constraints

output from NEXTTO
TITLE S1r
Coordination number distributions

Central particle: 1
Distance: 2.500A
Coordination number distributions
BARCHART
CURVES
  4 1

          0   0
          1   1171
          2   1810
          3   1019

Coordination numbers:
  1.962000

Appendix XVI. RMC parameters, two-fold coordination

S1r
0.0335   ! number density
  1.85   ! cut offs
0.3   ! maximum. move
0.05   ! r spacing
.false.   ! whether to use moveout option
0   ! number of configurations to collect
1000 0   ! step for printing, plotting off
60 60   ! Time limit, step for saving
0 1 0 0   ! No. of g(r), neutron, X-ray, EXAFS expts
s1_2.fq
1 500   ! Range of points
0   ! Constant to subtract
1   ! Coefficients
0.0025   ! Standard deviation
.false.   ! Renormalise
.true.   ! Offset
3   ! no. of coordination constraints
1 1 0 2.5 2 1 le-5
1 1 0 2.5 4 0 le-5
1 1 0 2.5 5 0 le-5
0   ! no. of average coordination constraints
.false.   ! whether to use a potential

56

Appendix XVII. Coordination number distribution, two-fold coordination

output from NEXTTO
TITLE
s1r
Coordination number distributions

Central particle: 1
Distance: 2.500A
Coordination number distributions
BARCHART
CURVES
  6 1

          0   39
          1   126
          2   3759
          3   38
          4   36
          5   2
Coordination numbers:
  1.978000

Appendix XVIII. RMC parameters, to disorder the ring molecules

s1ring
0.0335   ! number density
  1.85   ! cut offs
0.3   ! maximum move
0.05   ! r spacing
.false.   ! whether to use moveout option
0   ! number of configurations to collect
1000 0   ! step for printing, plotting off
60 60   ! Time limit, step for saving
0 0 0 0   ! No. of g(r), neutron, X-ray, EXAFS expts
1   ! no. of coordination constraints
1 1 0 2.5 2 1
1e-6   ! no. of average coordination constraints
0   ! whether to use a potential
.false.

Appendix XIX. RMC parameters, fitting data with ring molecule constraint

```
s1ring
0.0335                  ! number density
  1.85                  ! cut offs
0.3                     ! maximum move
0.05                    ! r spacing
.false.                 ! whether to use moveout option
0                       ! number of configurations to collect
1000 0                  ! step for printing, plotting off
60 60                   ! Time limit, step for saving
0 1 0 0                 ! No. of g(r), neutron, X-ray, EXAFS expts
s1_2.fq
1 500                   ! Range of points
1                       ! Constant to subtract
1                       ! Coefficients
0.0025                  ! Standard deviation
.false.                 ! Renormalise
.true.                  ! Offset
1                       ! no. of coordination constraints
1 1 0 2. 5 2 1 le-6
0                       ! no. of average coordination constraints
.false.                 ! whether to use a potential
```