RMC_POT: A Computer Code for Reverse Monte Carlo Modeling the Structure of Disordered Systems Containing Molecules of Arbitrary Complexity

Orsolya Gereben[a] and László Pusztai*[a]

An approach has been devised and tested for preserving the molecular dynamics molecular geometry taking into account energetic considerations during Reverse Monte Carlo (RMC) modeling. Instead of the commonly used fixed neighbor constraints, where molecules are held together by constraining distance ranges available for the specified atom pairs, here molecules are kept together via bond, angle, and dihedral potential energies. The scaled total potential energy contributes to the measure of the goodness-of-fit, thus, the atoms can be prevented from drifting apart. In some of the calculations (Lennard-Jones and Coulombic) nonbonding potentials were also applied. The algorithm was successfully tested for the X-ray structure factor-based structure study of liquid dimethyl trisulfide, for which material now significantly more sensible results have been obtained than during previous attempts via any earlier version of RMC modeling. It is envisaged that structural modeling of a large class of materials, primarily liquids and amorphous solids containing molecules of up to about 100 atoms, will make use of the new code in the near future. © 2012 Wiley Periodicals, Inc.

DOI: 10.1002/jcc.23058

Introduction

The Reverse Monte Carlo (RMC) method[1–3] is a commonly used technique for the structural modeling of liquids and amorphous materials based on diffraction (and/or EXAFS) data or (partial) radial distribution functions, (p)rdf [$g(r)$]. It enables one to build three-dimensional structural models that are consistent with each data set considered during the process. In the liquid state often molecules have to be handled, where a reasonable molecular geometry has to be preserved (i.e., the drifting apart of atoms has to be prevented). For this purpose, the neighbor list based "fixed neighbor constraints" construction (FNC)[4] has usually been used, where all the important atomic pair distances in a molecule are constrained, that is, they have to remain within a predefined range. Originally even the starting set of particle coordinates ("configuration") had to satisfy the FNC constraints, making the creation of such a configuration sometimes rather tiresome. Later,[5] especially as more often configurations from molecular dynamics (MD) simulations were used for starting RMC modeling, the RMC computer code was modified so that in the starting configuration intramolecular atomic pairs may be outside the FNC range. In this case, only moves that push connected intramolecular atomic pairs closer to the prescribed FNC range are accepted. If the FNC ranges are too narrow the molecules become rigid and it becomes almost impossible to move the atoms around. If, conversely, the ranges are too wide then molecules become rather distorted since all the distances inside the (wide) FNC range are equally acceptable (including the higher and lower ends of the ranges). That means that the larger the molecules are the more problematic the FNC-based approach becomes.

For the above reasons, a different approach is introduced here for keeping molecules together. It has been a long standing goal of RMC development to somehow combine RMC modeling with MD computer simulations, thus, integrating potential calculation into structural modeling; the new code, RMC_POT,[6] has been developed with this idea in mind. RMC_POT is based on the RMC++[5] software; now the possibility of applying nonbonded interatomic potentials [Lennard-Jones (LJ) and Coulombic interactions], as well as intramolecular bond stretching, angle bending and dihedral stretching potential functions is included.

A similar approach was used by Morita et al.,[7] where the intramolecular structure is constrained by molecular mechanics calculations using bonded interactions; only the intermolecular interactions are modeled by RMC, and no nonbonding potentials are used. Soper[8] used empirical potential structure refinement to adjust a starting interatomic potential energy function to produce the best possible agreement between simulated and measured structure factors. As preliminary to this work, it is also mentioned that the "hybrid RMC" scheme developed by Opletal et al.[9] also uses energetic considerations for accepting/rejecting moves.

The new approach was tested on dimethyl trisulfide (DMeTS). We have already conducted a combined MD+RMC study (the latter using FNC) for this material[10]: this is why this liquid seemed to be a good candidate for a methodological comparison.

O. Gereben, L. Pusztai
Institute for Solid State Physics and Optics, Wigner Research Centre for Physics, Hungarian Academy of Sciences, PO Box 49, H-1525 Budapest, Hungary
E-mail: pusztai.laszlo@wigner.mta.hu

Contract/grant sponsor: Hungarian National Basic Research Fund (OTKA); Contract/grant number: 083529.

In "Calculation of the Potential Energy" section, elements of the new algorithm are presented whereas "Simulation Details" section describes calculation details for liquid dimethyl sulfide. In "Results and Discussion" section, results for the test case are reported and compared; for completing the comparison, relevant MD and FNC-based RMC calculations are also shown. Finally, in "Conclusions" section, conclusions are drawn.

Calculation of the Potential Energy

RMC_POT[6] makes use of the logics and some particular elements introduced by the GROMACS MD software package[11] (for instance, the binding structure of the molecules is defined in the "topology" file[11]). This way, the best compatibility with GROMACS can be achieved, so that particle configurations, including description of the molecules, produced for GROMACS and RMC_POT are mutually interchangeable. This particular choice makes (arbitrarily) frequent switching between the two computational methods feasible during the process of structural modeling. Switching from a Monte Carlo type algorithm to an MD type one may be necessary generating significant displacements (rotation, translation) of whole molecules.

Nonbonded potentials are handled as pair potentials, consisting of dispersion and repulsion parts, making up the van der Waals term (presently only the LJ potential is implemented). Permanent charges act through the Coulombic term. Nonbonded interactions are calculated (up to a certain cutoff radius) according to

$$V_{\mathrm{LJ}}(r_{ij}) = 4\varepsilon_{ij}\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{N} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right)$$
$$V_{\mathrm{Coulomb}}(r_{ij}) = \frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{r_{ij}} \quad (1)$$

where ε_{ij} (kJ/mol) and σ_{ij} (Å) are the LJ parameters of the *ij* atom pair, whereas q_i and q_j are the partial charges placed on the atoms located at a distance r_{ij} from each other. ε_0 is the vacuum permittivity.

The essence of the RMC_POT algorithm is that bonded interactions are exploited for keeping molecules together, helped by nonbonding (LJ and Coulombic) interactions keeping meaningful molecular geometries. This principle is realized by the software according to the following consideration: if bonded atoms try to move away then the "goodness-of-fit" (χ^2) contribution defined for the bonding energy increases, thus decreasing the likelihood of accepting that particular move. The calculation of the bonded interactions follows the conventions set by the GROMACS MD suite.[11]

The bond stretching interaction between atoms *i* and *j* can be approximated as a harmonic potential:

$$V_b(r_{ij}) = \frac{1}{2}k_{ij}^{b}\left(r_{ij} - b_{ij}\right)^2 \quad (2)$$

where the k_{ij}^{b} is the force constant (kJ mol^{-1} nm^{-2}) and b_{ij} is the equilibrium distance (in nm; should be defined in the topology file, see the RMC_POT user guide for details[6]).

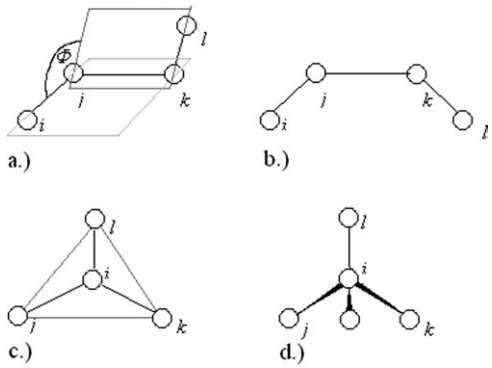

Figure 1. (a) Proper dihedral angle. Improper dihedral for a ring (b), for a planar group (c) and for a chirality center (d).

The harmonic angle bending potential for atoms *i*, *j*, and *k* (*j* is in the middle) is defined as:

$$V_a(\theta_{ijk}) = \frac{1}{2}k_{ijk}^{a}\left(\theta_{ijk} - \theta_{ijk}^{0}\right)^2 \quad (3)$$

where the k_{ijk}^{a} is a force constant (kJ mol^{-1} rad^{-2}) and θ^0 is the equilibrium angle (in radian; should be defined in the topology file).

Dihedral angles may be proper or improper. A proper dihedral angle is defined according to the IUPAC/IUB convention[12] as the angle between planes *ijk* and *jkl* (Fig. 1), with zero corresponding to the *cis* conformation.

Improper dihedrals are used to keep planar groups flat, or to prevent chiral groups to transform to their mirror image.

Three different definitions for a dihedral angle are possible: the periodic, harmonic, and so-called Ryckaert–Bellemans[13] (RB) dihedral formulation; the exact way the proper and improper dihedrals are calculated depends on the type of the force field.

Usually periodic dihedrals are used to handle proper dihedral interaction; the energetic contribution due to their variation can be calculated as:

$$V_{pd}(\Phi_{ijkl}) = k^{\phi}\left(1 + \cos\left(n\Phi_{ijkl} - \Phi^{s}\right)\right) \quad (4)$$

where k^{Φ} is the force constant (kJ/mol), *n* is the multiplicity, and Φ^{S} is the equilibrium angle in radian (should be defined in the topology file).

The energies related to improper dihedrals may be calculated according to the harmonic dihedral formulation:

$$V_{hd}(\xi_{ijkl}) = \frac{1}{2}k_{ijkl}^{\xi}\left(\xi_{ijkl} - \xi^{0}\right)^2 \quad (5)$$

where k^{ξ} is the force constant (kJ mol^{-1} rad^{-2}) and ξ^0 is the equilibrium angle value in radian.

For alkanes the RB [13] dihedral definition is often applied:

$$V_{\mathrm{RB}d}\left(\psi_{ijkl}\right) = \sum_{n=0}^{5} C_n(\cos(\psi_{ijkl}))^n \quad (6)$$

where $\psi_{ijkl} = \phi_{ijkl} - 180^\circ$ (with the value of 0 corresponding to the trans conformation) and C_n are the RB constants (kJ/mol).

The above potential energy terms give a contribution to the overall goodness-of-fit, χ^2. Two different ways have been implemented as to how the energy-related contribution might be best adjusted to the structure related one. According to the first route, separate potential energy and structure related "χ^2"-s are defined. The potential energy related χ^2_{pot} is evaluated first, in a way that is similar to the standard χ^2: if the new χ^2_{pot} is lower than it was in the previous step then the move is accepted; if otherwise then the move is accepted only with a probability exp $\{-(\chi^2_{pot,new} - \chi^2_{pot,old})/2\}$, satisfying the detailed balance condition. The move is then assessed on the basis of the standard (structure related) χ^2. According to the other route, the potential energy related terms contribute to the standard χ^2, together with the data series, so that acceptance/rejection of a move is a one-step process.

The chi-squared contributions coming from the bonded and nonbonded interactions are calculated as $V_i/\sigma_i{}^2$, where i refers to the potential component in question, and σ_i is the weighting factor for the given interaction. Whether this contribution is added to the standard χ^2 or set up a separate χ^2_{pot} (i.e., whether the first or the second approach, described above, is followed) depends on the particular case under investigation.

Simulation Details

The experimental X-ray data was measured at room temperature in the SPring-8 synchrotron radiation facility (Japan).[10] Here, only the formalism regarding the calculation of the X-ray data is given; note that the code is capable of using also neutron diffraction and EXAFS data as experimental data sets. Their calculation can be found in the description of earlier versions of our RMC++ code.[5,6]

Formalism of the structure-related function

In RMC_POT, total scattering structure factors (i.e., the experimentally determined information) are computed from the simulation box via the following route: (1) partial structure factors (psf), $A_{ij}(Q)$, are calculated via Fourier transformation from the partial radial distribution functions, $g_{ij}(r)$, according to

$$A_{ij}(Q) - 1 = \frac{4\pi\rho}{Q}\int_0^\infty r\left(g_{ij}(r) - 1\right)\sin(Qr)dr \qquad (7)$$

where r is the distance variable, ρ is the number density. (2) Psf's are then summed, by taking into account the c_i molar fractions and the Q-dependent atomic form factors $f_i(Q)$, to give the total scattering X-ray weighted structure factor, $F(Q)$.

$$F(Q) = \sum_{i,j=n} c_i c_j f_i(Q) f(Q)_j \left(A_{ij}(Q) - 1\right) \qquad (8)$$

MD details

1331 molecules were placed in cubic simulation cells during both RMC and MD simulations. The atomic number density was $6.30621 \cdot 10^{-2}$ Å^{-3}.

MD was used for creating initial configurations and *FNC* ranges for the traditional RMC. An MD simulation identical to that reported in Ref. [10] was carried out for the present project; the calculation was performed by using the GROMACS simulation package version 4.0.[11] The Berendsen thermostat[14] was applied for keeping the temperature constant (with temperature coupling time constant $\tau_T = 0.1$ ps) at $T =$ 293 K in the NVT ensemble. The leap-frog algorithm was used for integrating Newton's equations of motion, with a time step of $dt = 2$ fs in each case. The total energy reached its equilibrium value within less than 100 ps. MD configurations were collected 20 ps apart. The last 1500 ps of the 2000 ps trajectory was used for calculating the *prdf*'s, the same way as in our RMC code; the g_*rdf* software of the GROMACS package has been modified for this purpose.

Interatomic potentials

As for force field, the all atom OPLS-AA force field of Jorgensen et al.[15] was selected; this choice is related to the requirement that in RMC, each atom of the molecules has to be represented separately so that the calculation of the adequate structure factor may be performed. The OPLS force field uses effective pair interactions, including the 6–12 LJ potential for representing dispersion and repulsion terms and the Coulomb interaction [see above, eq. (1)]. LJ parameters and partial charges are given in Table 1. In our earlier, detailed study of DMeTS[10] "disulfide S" parameters have always been taken for the two outer sulfur atoms. As no OPLS-AA parameter is available for the middle S atom, different simulations were conducted using either the sulfide or the disulfide sulfur LJ parameters, in both cases with zero charge. A better agreement with the experimental data could be achieved using sulfide S for the middle sulfur and therefore this was the choice here, as well.

Table 1. Lennard-Jones parameters and partial charges for the atom types used in the simulations.

Name	σ(Å)	ε (kJ mol)	q (e)
S (sulfide for the middle S)	3.6	1.48532	0
S (disulfide)	3.55	1.046	−0.217
C (with disulfide)	3.5	0.276144	0.037
H	2.5	0.12552	0.06

The cutoffs for the LJ and short-range part of the Coulomb interactions were 11 and 9 Å, respectively. For the long-range part of the Coulomb interactions particle-mesh Ewald[16] summation, with interpolation order 4 and grid size of 0.12 nm, was applied. The relative strength of the electrostatic interaction at the cutoff is $\sim 10^{-5}$, thus providing an accuracy of $5 \cdot 10^{-3}$; this value is better than the accuracy of the LJ calculations.

The applied bond and angle and dihedral angle parameters are listed in Table 2. No SSS angle parameter is found in the OPLS-AA force field, so the experimental value of Donohue

Table 2. Equilibrium bond length, bond angle and RB dihedral parameters used in MD and RMC_POT simulations.

Bond	b_e (Å)	Source
CH	1.09	OPLS-AA
CS	1.81	OPLS-AA
SS	2.038	OPLS-AA
Angle	θ (°)	Source
HCH	107.8	OPLS-AA
HCS	109.5	OPLS-AA
CSS	103.7	OPLS-AA
SSS	104	Donohue experimental[14]

RB dihedral	C_0 (kJ mol^{-1})	C_1 (kJ mol^{-1})	C_2 (kJ mol^{-1})	C_3 (kJ mol^{-1})
HCSS	1.16734	3.50201	0.00000	−4.66935
CSSS	27.45332	10.70058	31.02018	−14.26744

and Schomaker[17] is adopted. Dihedral angles are calculated as RB dihedrals in the OPLS-AA framework. As no dihedral angle exists in the OPLS-AA force field for the CSSS section, the CSSC dihedral angle is used instead (note that the RB C_4 and C_5 coefficients are zero for both dihedrals and therefore they are not mentioned in table).

RMC calculations

Various RMC calculations, applying different sets of constraints, have been conducted, so that the capabilities of the new approach may become better visible (see Table 3 for a summary of the calculations). RMC simulations using potentials were carried out by the RMC_POT[6] 1.1 suite. Results were compared to the RMC simulations of our previous study[10] in which the so-called "*0p2w*" FNC ranges were applied ("*0p2w*" means that the bonded or nonbonded distance distributions were cut at the distance values—thus setting the boundaries of the FNC range—where the magnitude of the distribution was 20% that of the maximum value, see Table 4). In the potential-free RMC calculations, nine *FNC* constraints were constructed for defining molecules: three for bonds, four for 1–3 distances (angles), one for the nonbonding 1–4 C—S and one for the 1–5 C—C distance. In some of the calculations, constraints for cosine distributions of all possible (HCH, HCS, CSS, and SSS) bond angles were constructed; these distributions were set so that they match the average angle distributions of the MD simulation as closely as possible. A bin size of $dr = 0.1$ Å was used for the *prdf*-histogram and of $dcos(\theta) = 0.05$ for the cosine distributions of bond angles constraints. The cutoff distances in the RMC_POT calculations were $d_{CC} = 3.1$ Å, $d_{CS} = 2.86$ Å, $d_{CH} = 2.2$ Å, $d_{SS} = 3.0$ Å, $d_{SH} = 1.9$ Å, and $d_{HH} = 1.65$ Å.

Table 3. The applied data sets, constraints, and calculation methods in the different RMC calculations.

Simulations	*F*(*Q*)	6 *g*(*r*)	4 cos (θ)	FNC	Flexible molecule using potentials	NB	Separate χ^2 evaluation
FNC_fq	+	−	−	+	−	−	−
FNC_fq_gr	+	+	−	+	−	−	−
FNC_fq_gr_ang	+	+	+	+	−	−	−
Fm_fq_sep	+	−	−	−	+	−	+
Fm_fq_NB	+	−	−	−	+	+	+
Fm_fq	+	−	−	−	+	−	−
Fm_fq_gr_sep	+	+	−	−	+	−	+
Fm_fq_gr_NB	+	+	−	−	+	+	+
Fm_fq_gr	+	+	−	−	+	−	−

"Separate χ^2 calculation" means that χ^2_{pot} was first separately evaluated before the standard (structure-based) χ^2 evaluation.

Table 4. FNC ranges applied in some of the DMeTS RMC simulations.

FNC constraints	"*0p2w*" range (Å)
C—S bond	1.75–1.87
S—S bond	1.96–2.11
C—H bond	1.03–1.15
H—H nonbonding	1.65–1.88
C—S 1–3 nonbonding	2.87–3.18
C—S 1–4 nonbonding	3.47–4.23
S—S nonbonding	3.03–3.37
S—H nonbonding	2.28–2.54
C—C nonbonding	3.53–5.39

The six partial *g*(*r*)-s from MD were also included as "experimental" data (see Table 3 for an identification of calculations) their weight parameters were then set so that their initial contribution to the composite χ^2 would be one-tenth of that of the contribution of the structure factor; if cosine distributions of the bond angles were also constrained, then the similar ratio was set to 2.5%. For the flexible molecule runs, the contribution of the bonded interactions, either to the standard χ^2 or to χ^2_{pot}, depending on the actual calculation, was set to make the bond interaction contributions 10% of the *F*(*Q*) contribution. The same weighting was applied for the angular potential contribution whereas the weight of the dihedral potential was twice that (thus making the dihedral contribution smaller).

As indicated above, several different flexible molecule-containing RMC calculations were started from the same MD configuration (Table 3). Three standard, FNC-containing, RMC simulations reported earlier[10] are mentioned here for comparison (the run names beginning with "*FNC*"). The names of the flexible-molecule simulations begin with "*Fm*". "*fq*" in the name means that the simulation fitted the X-ray structure factor while "*gr*" refers to fitting the six partial *g*(*r*)'s from MD. "*ang*" means that the four cosine distributions of bond angles of the MD results were also fitted; the squared differences between theoretical and calculated distributions are calculated similarly to the other data sets.

Simulations *Fm_fq* and *Fm_fq_gr* applied "flexible molecule" constraints (potentials) which gave contribution to the standard χ^2 only; in the cases of calculations *Fm_fq_sep* and *Fm_fq_gr_sep* χ^2_{pot} was first separately evaluated. RMC runs *Fm_fq_NB* and *Fm_fq_gr_NB* utilized both flexible molecules and nonbonding interactions; here contributions from all potential related interactions composed χ^2_{pot}. As it was mentioned before, if χ^2_{pot} is calculated then it is evaluated before the standard "structure-related" χ^2, so only moves that have

already passed the potential-based criteria are evaluated further. RMC simulations performed this way would most certainly follow a different trajectory in configuration space from those where the potential-related χ^2 contributions are added to the standard χ^2.

Regardless how a particular simulation has been performed, χ^2 (with $\sigma = 1$) values were computed for the final configurations, individually for the *F*(*Q*) data set, the *prdf*-s, the cosine distributions, the bond-length, bond angle, dihedral angle (always with respect to the "experimental" values). LJ and Coulomb potential energies were also calculated.

As mentioned above, the RMC_POT code was developed from our earlier computer code RMC++,[5] and therefore, it has been written in standard, platform-independent C++. The code can be compiled for either consecutive or parallel execution. Parallelization is achieved using the Portable Operating System Interface (POSIX) thread standard, which is for shared memory multiprocessor computers, distributing the work where it is possible among the threads. The code is self-contained, only the POSIX interface has to be installed for the parallel version. The performance efficiency is defined as $E(p) = \frac{C_1}{pC_p}100\%$, where C_1 is the running time of the program using consecutive execution, C_p is the running time of using *p* threads, where each thread is executed by its on processor core. For a system of 114,244 atoms using an *F*(*Q*) data set with 190 points, coordination constraint and nonbonding potential, the efficiency of *E*(4)∼83% or *E*(8)∼90% was achieved. The code is available on the Internet[6] under the conditions of the GNU public license.

The RMC and RMC_POT calculations reported here usually took no more than a couple of hours on an Intel(R) Xeon(R) CPU E5345 2.33GHz 2*quad core processor based computer using four threads.

Results and Discussion

Before going into details, it is important to check whether the molecules were really kept together sufficiently and also, whether they also kept a meaningful shape during RMC simulations containing flexible molecules. It has therefore also been determined whether the final configurations of these simulations would satisfy the *FNC* constraints. All the atomic distances of the new flexible molecule RMC calculations fell within the full *FNC* range (i.e., within the original MD-determined distance range). Also, the number of distances outside the narrow "*0p2w*" *FNC* range was lower (13%), than for the starting MD configuration (17%) (see Ref. [10] for a description and Table 4 for the relevant distance ranges). This means that flexible molecular potential-based constraints are, indeed, able to hold molecules appropriately together. That is, the basic aim of the present methodological development has been achieved.

Assessment on the grounds of goodness-of-fit

First, let us compare the three flexible molecular RMC computations using only the experimental structure factor. As it can be seen in Table 5, simulations *Fm_fq_sep*, *Fm_fq_NB*, and *Fm_fq* all produced the same very good fit to the *F*(*Q*) ($\chi^2 \cong$ 0.15, see Fig. 2a). χ^2 calculated via comparison of the partial *rdf*-s of the final configurations to the average MD partial *rdf*-s are very similar for the three models, so are the χ^2 resulting from the comparison of the four different (CSS, SSS, HCH, and HCS) cosine distributions of bond angles. Comparing the values to those obtained for the traditional *FNC*-using *RMC* simulation *FNC_fq*, the *F*(*Q*) fits are similarly good. However, the potential-based runs with flexible molecules produced configurations with much more MD-like (and physically much more feasible) partial *g*(*r*)-*s* and cosine distributions of bond angles (total χ^2 ∼60), than calculation *FNC_fq* (total $\chi^2 = 1408.5$). The most problematic *prdf* in run *FNC_fq* was the S-H partial *g*(*r*),

Table 5. Chi-square contributions of the various data sets/constraint elements (some of them determined not during, but after the simulation; each value was calculated with $\sigma = 1$).

	χ^2			
Simulations	*F*(*Q*)	Total *g*(*r*)	TotalS *cos* (*θ*)	Total
FNC_fq	0.10	1269.33	139.07	1408.50
FNC_fq_gr	1.06	0.05	11.45	12.56
FNC_fq_gr_ang	1.13	0.04	0.00	1.17
Fm_fq_sep	0.15	25.32	31.54	57.02
Fm_fq_NB	0.14	34.90	29.43	64.47
Fm_fq	0.15	27.17	35.63	62.96
Fm_fq_gr_sep	1.01	0.06	1.78	2.85
Fm_fq_gr_NB	1.12	0.05	0.81	1.98
Fm_fq_gr	1.03	0.05	1.31	2.40

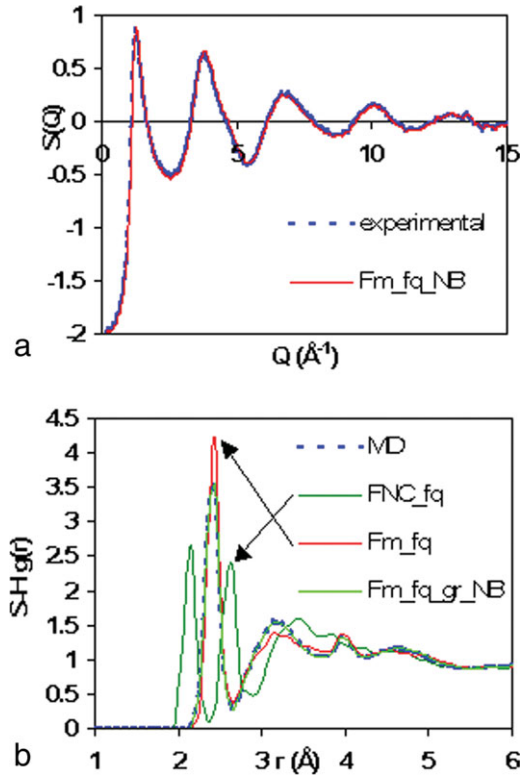

Figure 2. (a) Comparison of the calculated total structure factor to the experimental *F*(*Q*) for the *Fm_fq_NB* model. (b) The S-H *prdf* for simulations *FNC_fq*, *Fm_fq_NB*, *Fm_fq_gr_NB*, as compared to the corresponding MD result. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

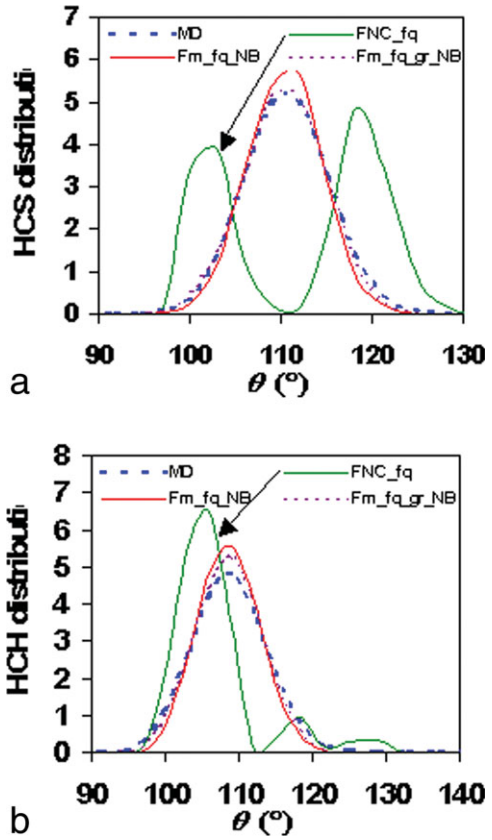

Figure 3. Comparison of the (a) HCS and (b) HCH cosine distributions of bond angles from RMC simulations *FNC_fq*, *Fm_fq_NB*, and *Fm_fq_gr_NB* to the corresponding average MD data (note that on the *x*-axis now θ is displayed, instead of *cos*(*θ*)). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

where the first peak resulting from the FNC-based calculation completely split into two (Fig. 2b). As for the cosine distributions, the largest discrepancies can be found in terms of the hydrogen-containing HCS and HCH angles (Figs. 3a and 3b). Note that for the potential-based RMC simulations only a small difference in terms of the peak height is visible, as compared to the MD average (Figs. 3a and 3b).

If the partial *rdf*-s were also fitted together with *F*(*Q*) then the flexible molecule simulations (*Fm_fq_gr*, *Fm_fq_gr_sep*, and *Fm_fq_gr_NB*) produced also much better results (total $\chi^2 = 2.36$–2.85) than the corresponding traditional calculation *FNC_fq_gr* (total $\chi^2 = 12.56$) (Table 5). Even the problematic S—H partial became almost identical to the original MD *prdf* (Fig. 2b), and the HCH and HCS angular distributions became almost identical to the average MD cosine distributions (Figs. 3a and 3b). It appears that RMC simulations using 'potential based' flexible molecules can reproduce the *rdf*-s and cosine distribution of bond angles inherently much better; a possible reason is that in the new algorithm, bonds and angles around the maximum peak height of the corresponding distributions are preferred.

Assessment based on energetic considerations

Now consider the potential energies of the different simulations. The bonded interaction energies were the tool for keeping the molecules together. It is obvious from the inset of Figure 4 that the total bonded energy (calculated after the simulations for the traditional *FNC*-based runs) is higher than that of the starting MD simulation for all the normal RMC-*FNC* runs, especially for the *FNC_fq* simulation.

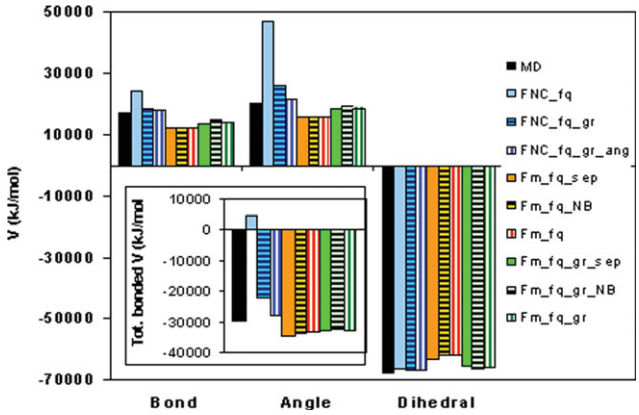

Figure 4. Energies of the bonded interactions for all the simulations. Main chart: bonding, angular and dihedral contributions to the total bonding potential energy. Inset: total bonding energy.

For each potential-based flexible molecule RMC calculation the total bonded energy became lower than even the original MD energy. If we decompose the total energy in to bonding, angular and dihedral components, shown on the main chart of Figure 4, we observe the same tendency for the bonding and angular potential energies. The dihedral potential energies (and the resulting dihedral distributions, not shown) of all the simulations are similar to each other, as well as to the same quantity calculated for the starting MD configuration.

It is interesting to notice that the traditional *FNC_fq* simulation using only the X-ray total scattering structure factor, that produced such a good fit in *Q*-space, resulted in an energetically very unfavorable state; this had already been indicated by the unrealistic S-H partial *g*(*r*), as well as by the angular distributions.

The nonbonded energies were only minimized in calculations *Fm_fq_NB* and *Fm_fq_gr_NB*; they were, however, computed for the final configurations of each RMC simulation. The total nonbonded energy is shown in the inset of Figure 5; its

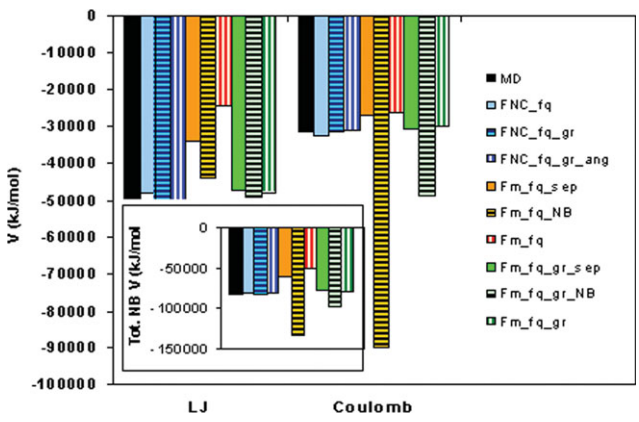

Figure 5. Nonbonded interaction energies for all the simulations. Main chart: LJ and Coulombic contributions to the total nonbonding interaction potential energy. Inset: total nonbonded energies.

components are depicted in the main chart of Figure 5. For the *FNC*-using simulations the total nonbonded energy remained almost the same as for the starting MD simulations. In details, the Coulomb part decreased and the LJ part increased somewhat. For potential-based RMC calculations that use only the $F(Q)$ as input, the nonbonded energy has decreased strongly only for the *Fm_fq_NB* simulation. Here, too, the decrease can be originated to the decreasing Coulombic energy; the LJ energy has increased somewhat as compared to the starting MD.

For potential-based RMC calculations where both the $F(Q)$ and the full set of (six) partial $g(r)$-s were applied as input structural information, there is a decrease in terms of the total nonbonded energy for *Fm_fq_gr_NB*, although much smaller than detected for simulation *Fm_gr_NB*. The decrease here also originates in the Coulomb part. In calculations *Fm_fq_gr* and *Fm_fq_gr_sep* both the total nonbonding potential energy and its components remained nearly the same as for the MD. The reason why the decrease came mainly from the Coulombic contribution is most probably because the contribution of the Coulomb potential to the nonbonded potential was much larger than that of the LJ part; that is, the decrease of the Coulomb term had to be the driving force of the minimization of the nonbonding potential energy.

Conclusions

An efficient algorithm for keeping molecules together during RMC calculations has been devised and tested. Flexible molecules were kept together by bonded interaction potentials, defined for bond lengths, bond angles and dihedral angles. In addition, the nonbonded potential energy (LJ and Coulomb terms) can also be minimized.

The new approach was tested on DMeTS, which was previously investigated by traditional, *FNC*-using, RMC. It could be shown that the potential-based flexible molecule approach produces better, energetically more reliable simulation results while using less additional constraints. This approach, and the computer code RMC_POT that realizes it, is expected to be especially suitable for modeling larger molecules, where the application of the traditional approach that uses the *FNC* concept would be extremely cumbersome, or even impossible.

As far as moving molecules around efficiently, thanks to the uniform treatment of potential functions and molecular topologies in the GROMACS MD package and RMC_POT, switching between RMC and MD simulations is easier than ever. That is, during modeling the structure of a given molecular material, proper exploration of the configuration space can now be carried out.

Keywords: liquid structure · molecular liquids · computer modeling · Reverse Monte Carlo

How to cite this article: O. Gereben, L. Pusztai, *J. Comput. Chem.* **2012,** *33,* 2285–2291. DOI: 10.1002/jcc.23058

[1] R. L. McGreevy, L. Pusztai, *Mol. Simul.* **1988**, *1,* 359.
[2] L. Pusztai, *J. Non-Cryst. Sol.* **1998**, *88,* 227.
[3] R. L. McGreevy, *J. Phys. Cond. Matter* **2001**, *13,* R877.
[4] L. Pusztai, R. L. McGreevy, NFL Studsvik Annual Report for 1996, **1997**, OTH 21.
[5] O. Gereben, P. Jóvári, L. Temleitner, L. Pusztai, *J. Optoelectron. Adv. Mater.* **2007**, *9,* 3021.
[6] RMC_POT. Available at: http://www.szfki.hu/~nphys/rmc++/opening.html. Accessed on June 29, 2012.
[7] H. Morita, S. Kohara, T. Usuki, *J. Mol. Liq.* **2009**, *147,* 182.
[8] (a) A. K. Soper, *Chem. Phys.* **1996**, *202,* 295; (b) A. K. Soper, *Mol. Phys,* **2001**, *99,* 1503; (c) A. K. Soper, *Phys. Rev. B,* **2005**, *72,* 104204.
[9] (a) G. Opletal, T. Petersen, B. O'Malley, I. Snook, D. G. McCulloch, N. A. Marks, I. Yarovsky, *Mol. Simul.* **2002**, *28,* 927; (b) G. Opletal, T. C. Petersen, B. O'Malley, I. K. Snook, D. G. McCulloch, I. Yarovsky, *Comput. Phys. Comm.* **2008**, *178,* 777.
[10] O. Gereben, S. Kohara, L. Pusztai, *J. Mol. Liq.* **2012**, *169,* 63.
[11] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, *J. Chem. Theory Comput.* **2008**, *4,* 435; Available at: http://www.gromacs.org. Accessed on June 29, 2012.
[12] Available at: http://www.chem.qmul.ac.uk/iupac/iupac.html. Accessed on June 29, 2012.
[13] J. P. Ryckaert, A. Bellemans, *Faraday Disc. Chem. Soc.* **1978**, *66,* 95.
[14] H. J. C. Berendsen, J. P. M. Postma, A. DiNola, J. R. Haak, *J. Chem. Phys.* **1984**, *81,* 3684.
[15] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, *J. Am. Chem. Soc.* **1996**, *118,* 11225.
[16] U. Essman, L. Perela, M. L. Berkowitz, T. Darden, H. Lee, K. G. Pedersen, *J. Chem. Phys.* **1995**, *103,* 8577
[17] J. Donohue, V. J. Schomaker, *Chem. Phys.* **1948**, *16,* 92.

Received: 29 March 2012
Revised: 7 June 2012
Accepted: 8 June 2012
Published online on 10 July 2012