Accelerated simulations of aromatic polymers: application to polyether ether ketone (PEEK)

Richard J. Broadbent, James S. Spencer, Arash A. Mostofi & Adrian P. Sutton

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Accelerated simulations of aromatic polymers: application to polyether ether ketone (PEEK)

Richard J. Broadbent[a,b], James S. Spencer[a,b,c], Arash A. Mostofi[a,b,c] and Adrian P. Sutton[a,b,*]

aDepartment of Physics, Imperial College London, London, UK; bThe Thomas Young Centre for Theory and Simulation of Materials, Imperial College London, London, UK; cDepartment of Materials, Imperial College London, London, UK

For aromatic polymers, the out-of-plane oscillations of aromatic groups limit the maximum accessible time step in a molecular dynamics simulation. We present a systematic approach to removing such high-frequency oscillations from planar groups along aromatic polymer backbones, while preserving the dynamical properties of the system. We consider, as an example, the industrially important polymer, polyether ether ketone (PEEK), and show that this coarse graining technique maintains excellent agreement with the fully flexible all-atom and all-atom rigid bond models whilst allowing the time step to increase fivefold to 5 fs.

Keywords: coarse graining; rigid body; molecular dynamics; PEEK; constrained

1. Introduction

The dynamics of polymers span a wide range of time scales. Carbon–hydrogen covalent bonds have a vibration period of ∼10 fs [1], while macromolecules such as proteins undergo conformational changes over microseconds [2]. In practice, the maximum time scale accessible to molecular dynamics (MD) is limited by the time step required to capture the fastest degree of freedom in the system. It is desirable, therefore, to include only the degrees of freedom that are physically relevant to the problem under consideration.

In polymers, the fastest oscillations are usually associated with covalent bonds between heavy atoms (e.g., carbon, nitrogen, oxygen) and hydrogen. However, conformational changes are always associated with considerably longer time scales. Consequently, these fastest oscillations are often suppressed by applying constraints to fix bond lengths at their equilibrium values. This is usually accomplished by the SHAKE [3,4] or LINCS [5,6] constraint algorithms. Often, it is desirable to go further and to constrain other bond lengths and angles, enabling time steps of up to 2 fs [7]. While this approach can result in an increase in the duration of time that may be simulated, parallel scalability and algorithm stability can be affected when the constraints are highly coupled [6].

An alternative approach to applying constraints is the use of generalised coordinates representing rigid multibody elements [8], whereby atoms are grouped into rigid dynamical units. The removal of fast degrees of freedom may also be achieved through the introduction of 'virtual sites' for the hydrogen atoms, whereby the position of each hydrogen atom is defined by the positions of nearby heavy atoms, which can allow time steps of up to 7 fs [1]. A drawback of this approach is that the moment of inertia tensor of the molecule is not conserved [1] (e.g., the principal values for benzene can have errors greater than 10%) which may affect the large scale dynamics of the system.

In this paper, we present a coarse-graining approach that addresses this limitation for the case of aromatic groups in polymers. Each aromatic group is mapped onto a rigid triangle, defined by three vertices, *in a manner that conserves key dynamical quantities*, namely, the total mass, the centre of mass, and the moment of inertia tensor. As a result, this triangle has the same dynamical response as a rigid, planar aromatic group. The fast degrees of freedom associated with bond vibrations and out-of-plane buckling modes *within* each group are removed. At the same time, the slower degrees of freedom, which are responsible for conformational changes, associated with variations in bond and torsion angles *between* adjacent groups are retained. The fixed relationship between the coordinates of the vertices of the triangle and the true atomic positions associated with the aromatic group uniquely determine the force and torque on the triangle for a given set of positions and a given all-atom (AA) force field. Our approach is equivalent to coupled rigid body dynamics [9], but critically does not require altering the core integration routine of an existing MD code. Our coarse-grained (CG) method is equally applicable to aromatic groups that form the polymer backbone as well as those present in side chains.

We have tested our approach across a wide range of temperatures and system sizes for polyether ether ketone (PEEK), a widely used industrial polymer [10–12].

---

*Corresponding author. Email: a.sutton@imperial.ac.uk

Structural and dynamical properties are found to be in excellent agreement with both AA and AA rigid bond (AA-RB) simulations, and the increased MD time step ($\sim$5 fs) enabled by our method results in a fivefold increase in the time scale of a given simulation as compared to AA approaches.

The remainder of the paper is organised as follows. An introduction into the principles of rigid dynamics and how these can be applied to aromatic polymers to obtain a CG representation of planar units. The application to PEEK is used to illustrate this approach. A comparison of the results obtained from AA, AA-RB, and CG simulations of PEEK for a variety of polymer lengths and temperatures. Finally, an analysis of the coarse-graining technique and its applicability to other aromatic polymers.

2. Method

Considerable computational savings may be realised by removing the fast degrees of freedom normally associated with planar units in molecules, such as aromatic groups. Any rigid two-dimensional object has six independent quantities that govern its dynamics. For a planar object in the x–y plane, where x and y denote Cartesian axes, these are the position coordinates associated with the centre of mass, $\mathbf{R}^0 = (X^0, Y^0)$, the total mass M_T, and the three independent components of the moment of inertia tensor $\mathbf{I}$, which we denote as I_{xx}, I_{yy}, and I_{xy}. Any two rigid bodies having these properties in common will exhibit the same dynamical response when subjected to the same set of forces. Our coarse-graining approach, described below, explicitly conserves these properties.

2.1. Two-dimensional rigid bodies

Consider a two-dimensional rigid object composed of N atoms.[1] Let the x- and y-axes be aligned with the principal axes of $\mathbf{I}$ and let the origin be at $\mathbf{R}^0$, without loss of generality. Therefore,

$$\mathbf{R}^0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{1}$$

$$M_T = \sum_{i=1}^{N} m_i, \tag{2}$$

$$\mathbf{I} = \begin{pmatrix} \sum_{i=1}^{N} m_i y_i^2 & 0 \\ 0 & \sum_{i=1}^{N} m_i x_i^2 \end{pmatrix}, \tag{3}$$

where the ith atom has position coordinates (x_i, y_i) and mass m_i. In order to evolve the positions of this rigid body using MD, a naive approach might be to apply a set of constraints among the atoms such that only the rigid body degrees of freedom remain. However, this introduces $3(N-2)$ constraints, the enforcement of which becomes unstable as N
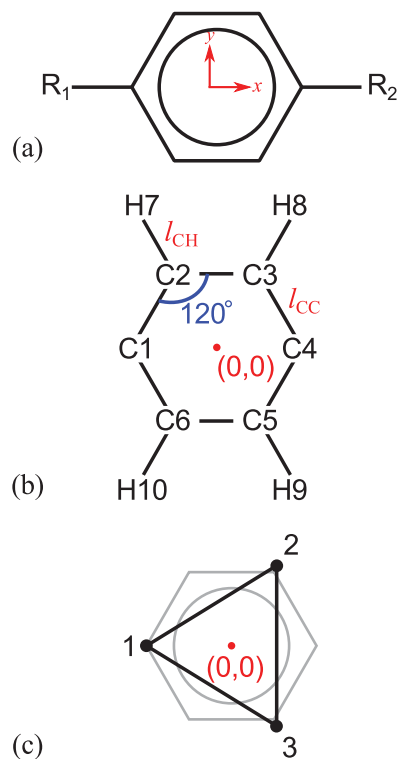

Figure 1. A schematic demonstration of the coarse-graining procedure of a para-substituted aromatic group in an organic molecule. (a) The coordinate system and chemical formula for the aromatic group with the remainder of the aromatic molecule indicated by R_1 and R_2. (b) The relevant bond angles and lengths where l_{CC} and l_{CH} are the equilibrium lengths of the carbon–carbon and carbon–hydrogen bonds. The atoms in this AA representation are labelled for reference. (c) The resulting triangle of the CG representation with the vertices labelled 1, 2, and 3 to match the text. The underlying molecule in grey is the AA group being replaced. All three figures are to scale using the parameters from the OPLS-AA force field [13]. The positions and masses of the triangle vertices in (c) are given in Table 1. The orientations of the triangles in the PEEK monomer are arbitrary as it would have been equally possible to place vertex 1 at the position of C4 at the beginning of the coarse-graining procedure. Figure 2 shows the orientation used in all CG simulations.

increases owing to the coupled nature of the constraints. Furthermore, constraints on light atoms such as hydrogen become unstable as the time step increases [1]. It is desirable, therefore, to develop a coarse-graining scheme that exhibits stability at large time steps while preserving dynamical properties.

2.2. Coarse-grained representation of aromatic polymers

We consider the case of a para-substituted aromatic group in a polymer backbone, as shown schematically in Figure 1. An example of a polymer that exhibits this structural feature is the macromolecule PEEK, as shown in Figure 2(a). Assuming planarity, rigidity, and symmetry, and using the
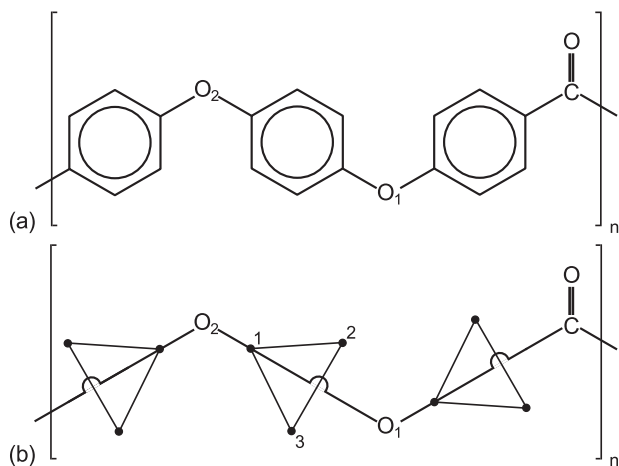
Figure 2. The chemical structure of PEEK, and the corresponding CG representation. The suffixes on the oxygen atoms are for identification purposes only – they do not indicate molecular or atomic oxygen. The constraints used for the middle triangle connect the pairs: O_2–1, 1–2, 1–3, 2–3, and 1–O_1. The constraints for the other triangles are analogous to this, and the oxygen in the ketone group is connected with a single constraint to the carbon atom.

notation of Equations (1)–(3), the AA representation of the aromatic unit (Figure 1(b)) is characterised by

$$\mathbf{R}^0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{4}$$

$$M_T = 6m_C + 4m_H, \tag{5}$$

$$I_{xx} = 3m_C l_{CC}^2 + 3m_H \left(l_{CC} + l_{CH} \right)^2, \tag{6}$$

$$I_{yy} = 3m_C l_{CC}^2 + m_H \left(l_{CC} + l_{CH} \right)^2, \tag{7}$$

$$I_{xy} = 0, \tag{8}$$

where m_C and m_H are the masses of a carbon atom and a hydrogen atom, respectively, and l_{CC} and l_{CH} are the lengths of a carbon–carbon and carbon–hydrogen bonds in the unit, respectively. The choice of coordinates in Figure 1(a) is convenient as these axes are the eigenbasis for $\mathbf{I}$, as can be seen from Equation (8).

Table 1. The positions (X_α, Y_α) and masses M_α of the vertices α of a triangle with the same dynamics as a rigid aromatic group along the backbone of PEEK.

Vertex α	X_α (nm)	Y_α (nm)	M_α (g mol^{-1})
1	−0.140	0.000	25.871
2	0.072	0.133	25.114
3	0.072	−0.133	25.114

The six independent quantities expressed in Equations (4)–(8) can be conserved exactly by a CG representation composed of three point masses at the vertices of a triangle (Figure 1(c)). These vertices, which we label $\alpha \in \{1, 2, 3\}$, are associated with *nine* parameters: position coordinates (X_α, Y_α) and masses M_α. However, since there are only *six* independent constraints embodied in Equations (4)–(8), there is no unique solution.

Further constraints are added by the requirement that the group be connected via distance constraints to the rest of the organic molecule, which leads to the convenient choice that one of the vertices of the triangle should coincide with the position of an atom connecting the aromatic group to the rest of the polymer. We place vertex 1, therefore, directly at the position of the carbon atom labelled C1 in Figure 1(b), thus specifying $X_1 = -l_{CC}$ and $Y_1 = 0$, reducing the number of free parameters associated with the vertices of the triangle to seven.

Furthermore, by symmetry, $X_3 = X_2$, $Y_3 = -Y_2$, and $M_3 = M_2$, thus simultaneously satisfying Equation (8) and the requirement that $Y^0 = 0$ (Equation (4)), while also reducing the number of free variables associated with the CG representation to four:

$$X^0 = \frac{-M_1 l_{CC} + 2M_2 X_2}{M_T}, \tag{9}$$

$$M_T = M_1 + 2M_2, \tag{10}$$

$$I_{xx} = 2Y_2^2 M_2, \tag{11}$$

$$I_{yy} = l_{CC}^2 M_1 + 2X_2^2 M_2. \tag{12}$$

Solving these equations for M_1, M_2, X_2, and Y_2 in terms of the known AA quantities (Equations (4)–(8)), we find that

$$M_1 = \frac{M_T I_{yy}}{I_{yy} + l_{CC}^2 M_T}, \tag{13}$$

$$M_2 = \frac{1}{2}(M_T - M_1), \tag{14}$$

$$X_2 = l_{CC} \left(\frac{M_T}{2M_2} - 1 \right), \tag{15}$$

$$Y_2 = \sqrt{\frac{I_{xx}}{2M_2}}. \tag{16}$$

Taking equilibrium parameters for the AA representation from the OPLS-AA force-field [13] results in the positions and masses associated with the CG representation given in Table 1.

The CG representation shown in Figure 1(c) and defined by the parameters given in Table 1 has identical dynamical properties to the original rigid AA aromatic group of Figure 1(b). However, in order to use the CG representation

in an MD simulation of a polymer such as PEEK, it is also necessary to constrain the 'bond lengths' associated with the CG representation. For the central aromatic group in the PEEK monomer, as shown in Figure 2(b), these constraints are denoted as 1–2, 1–3, 2–3, 1–O_1, and 1–O_2, and are enforced using the SHAKE [3,4] algorithm.

2.3. Forces on rigid groups

The procedure set out in the previous subsection gives a CG representation composed of point masses that define a set of rigid, planar triangles with the same centre of mass, total mass, and moment of inertia as a rigid, planar AA representation. It is also necessary to ensure that the CG triangles experience the same force and torque of interaction with other objects as the groups they replace. This is achieved via the concept of 'virtual sites' [14], in a manner similar to the generalised coordinate approach of POEMS [8].

The virtual sites are the positions of the atoms in the AA representation of the aromatic group, and their position coordinates are defined in terms of their fixed relationship with the vertices of the triangle in the CG representation. In the case of rigid planar groups, the transformation between AA and CG representations is a simple linear mapping (more complex mappings are possible [1,14] but are not required for our CG approach). Given the vertices of a triangle at $\mathbf{r}_1$, $\mathbf{r}_2$, and $\mathbf{r}_3$, the position of an arbitrary point ($\mathbf{Q}$) in the plane defined by the triangle is given by

$$\mathbf{Q} = (1 - a - b)\mathbf{r}_1 + a\mathbf{r}_2 + b\mathbf{r}_3, \tag{17}$$

for some values of a and b that can be calculated for each virtual site once and for all using simple vector geometry. Performing this mapping on a set of (a, b) pairs derived from the equilibrium atomic coordinates, the positions of the atoms in the AA representation may be calculated quickly and easily from the positions of the CG masses at each time step.

Given an AA force field, the forces ($\mathbf{f}_1$, $\mathbf{f}_2$, $\mathbf{f}_3$) on the triangle vertices due to a force ($\mathbf{G}$) acting on an atomic site are found by taking the derivative of Equation (17):

$$\mathbf{f}_1 = (1 - a - b)\mathbf{G}, \tag{18}$$

$$\mathbf{f}_2 = a\mathbf{G}, \tag{19}$$

$$\mathbf{f}_3 = b\mathbf{G}. \tag{20}$$

Thus, the task of generating a separate force field for the CG representation of the molecule is avoided, and standard AA force fields which have been developed and tested for the same macromolecule can be used without alteration. In the event that a complete force field for the desired molecule is unavailable the reduction in the number of degrees of freedom enabled by our method correspondingly reduces the number of terms that need to be parametrised. Furthermore, the automatically generated atomic coordinates can be used in conjunction with a variety of existing tools for analysing MD trajectories.

3. Results

The CG representation of PEEK was compared with AA and AA-RB simulations for four different lengths of polymer comprising 4, 8, 12, and 16 monomers (the monomer unit is shown in Figure 2). The molecules were terminated with a hydrogen atom on the left and a phenyl ring group after the ketone on the right. The chains, therefore, consisted of 13, 25, 37, and 49 aromatic groups, respectively.

All simulations were conducted using the GROMACS [6,7,15–17] MD suite (v4.5.5 with double precision). The polymers were modelled using the OPLS-AA force field [13], with 1.1 nm cut-off radius, smoothly tapered in the final 0.05 nm. The leapfrog integration algorithm and, for simulations carried out in a canonical ensemble, a stochastic velocity rescaling thermostat [18] were used. For the CG and AA-RB simulations, constraints were enforced to a fractional tolerance of 1×10^{-10} with the SHAKE [3,4] algorithm. To ensure adequate statistical sampling, for each system, 16 evenly spaced, uncorrelated configurations were taken from a 2-ns simulation in a canonical ensemble at 1400 K and used as the initial starting points for further simulations. These initial configurations were then simulated at 300, 500, 700, 1000, and 1400 K. First, the polymers were equilibrated in a canonical ensemble for 1 ns, followed by a further 1 ns in a micro-canonical ensemble. Production runs were then carried out in a micro-canonical ensemble with samples taken every 16.8 ps over a total duration of 5 ns for the 4- and 8-monomer conformations, and over a duration of 10 ns for the 12- and 16-monomer conformations. The sampling interval was chosen because it was the smallest interval for which the derived properties for each polymer were unaffected by doubling the interval. All properties presented below were averaged over the 16 independent simulations of each polymer molecule and the resulting trajectories were analysed using the MDAnalysis tool kit [19].

We first consider the 8-monomer polymer. In Figure 3, we compare the energy drift at 300 K for AA, AA-RB, and CG simulations as a function of MD integration time step. For the CG simulations, the SHAKE algorithm was unable to converge for time steps larger than 7 fs (data not shown), while for the AA simulations, time steps larger than 1.25 fs resulted in unstable dynamics (the vibration period of carbon–hydrogen bonds in the system is approximately 11 fs).

The fractional tolerance of 10^{-10} used with the SHAKE algorithm for all constraints renders the integration algorithm almost perfectly reversible. With the small time steps
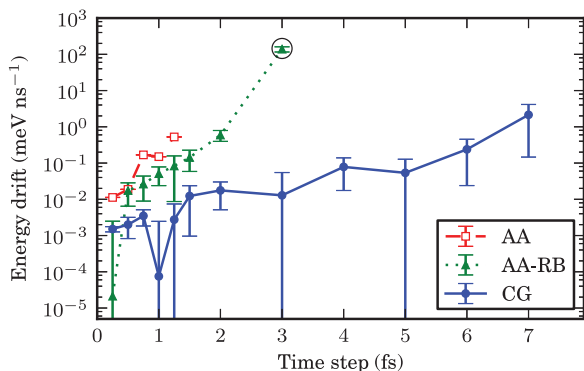
Figure 3. The absolute drift in total energy averaged over 16 independent 5-ns micro-canonical simulations of an 8-monomer polymer at 300 K for AA, AA-RB, and CG representations. As can be seen, with the exception of the 3-fs time-step AA-RB simulations (circled), all simulations showed excellent energy conservation.
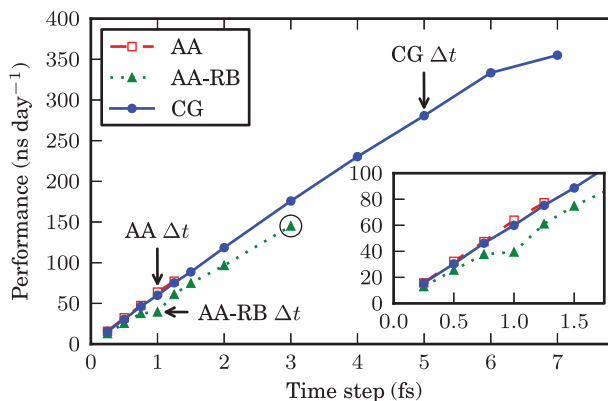

Figure 4. The performance averaged over 16 independent 5-ns micro-canonical simulations of an 8-monomer molecule at 300 K for AA, AA-RB, and CG representations. The AA simulation becomes unstable with a time step of 1.5 fs, and the energy drift of the AA-RB simulation with a 3-fs time step is too large. The arrows indicate the time steps used in the rest of the paper. The simulations were performed on one core of an Intel Xeon E5-2650.

where the AA and AA-RB simulations remain stable, the energy drift of the CG simulations is significantly smaller. At larger time steps, where the AA and AA-RB simulations become unstable, the energy drift of the CG simulations becomes comparable to that of the stable AA simulations. The energy drift of the CG simulations increases as the constraint tolerance is increased, as expected. Nevertheless, even when the fractional constraint tolerance was increased by three orders of magnitude to 10^{-7}, the CG simulations (not shown) remained in excellent agreement with the results shown in this section. For example, the energy drift with a 5-fs time step and a constraint tolerance of 10^{-7} was -31 ± 1 meV ns^{-1}. To compare this with other simulations using constraints, this energy drift is less than 25% of the drift per degree of freedom obtained for a box of 820 simple point charge (SPC) water molecules using analytic constraints, with a 4-fs time step and increased hydrogen masses to reduce the energy drift [1]. Therefore, we suggest that a SHAKE tolerance in the range $[10^{-10}, 10^{-7}]$ is suitable for most users.

To test the stability of the CG procedure over a long run we ran a simulation of a single 16-monomer polymer in the micro-canonical ensemble for 200 ns at 499.2 $\pm$ 0.3 K, with a time step of 5 fs. Throughout the 40 million MD time steps of the run, the fractional tolerance of 10^{-10} used with the SHAKE algorithm to enforce the constraints was maintained. This confirms the stability of the CG procedure throughout long simulations.

The LINCS algorithm was tested for a wide range of parameters (expansion order {4, 6, 8, 10, 12, 14, 16, 18, 20}, iterations {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}) with time steps of {1, 2, 3, 4, 5} fs. For each combination of parameters, 16 starting configurations were taken from CG simulations, successfully completed using SHAKE with the same time step. However, none of the 7200 simulations attempted

remained stable for more than four MD steps. This was caused by the highly coupled nature of the constraints which produce eigenvalues too large for the expansion used in LINCS. Therefore, the SHAKE algorithm was used for all constraints throughout this research and no further simulations using the LINCS algorithm were conducted.

The AA-RB simulations using a 3-fs time step (circled in Figure 3) result in an energy drift of 140 ± 20 meV ns^{-1}. This is 10 times the equipartition thermal energy of the system ($k_B T/2 \approx 13$ meV). It is caused by the rapid movement of the hydrogen atoms which cannot be accurately reproduced with this time step. At 300 K, a time step of 2 fs provides good energy conservation and stable dynamics. However, at 1000 K and above, the rapid movement of the hydrogen atoms led to failures by SHAKE to converge, or it produced erroneous constraint forces and concomitant spurious large changes in the total energy of the system. Both the SHAKE convergence failures and spurious energy changes are absent with a 1-fs time step in the AA-RB simulations.

Figure 4 shows the computational performance of simulations of the 8-monomer polymer at 300 K, as a function of the integration time step. It can be seen that for the same time step, CG simulations have performance comparable to AA simulations, and greater than AA-RB simulations. But the real advantage of the CG representation is that a larger time step can be used than with both the AA and AA-RB representations without sacrificing stability or energy drift. This performance enhancement is limited only by the maximum time step that can be used before the SHAKE algorithm fails to converge in a reasonable number (250) of iterations. At higher temperatures, the largest time step that can be used is reduced. For example, at 1400 K, a time step of less than 6 fs must be used (data not shown). For

all the simulations that follow, we used a 5-fs time step for CG, and a 1-fs time step for AA and AA-RB. With these parameters it can be seen in Figure 4 that the performance increase of our CG approach over the AA approach is a factor of five.

The instantaneous radius of gyration, R_g, is a key property of the molecular structure and is defined by

$$R_g^2 = \sum_{i=1}^{N} \left(\frac{\mu_i}{\mu_T} (\rho_i - \rho_c) \right)^2, \quad (21)$$

where N is the number of sites, μ_i and ρ_i are the mass and position, respectively, of site i, $\mu_T = \sum_i \mu_i$ is the total mass of the molecule, and ρ_c is its centre of mass. The term 'site' is used generically to refer to either an AA or a CG representation.

In Figure 5, we report the time- and ensemble-averaged radius of gyration $\langle R_g \rangle$ as a function of temperature for polymers ranging in length from 4 to 16 monomers. The time average is performed over the entire production run for each simulation, and the ensemble average is over 16 independent simulations for each polymer length. As can be seen from the figure, the CG simulations are in very good agreement with both the AA and AA-RB results.

In order to assess the similarity in the dynamics of the molecules in the three types of simulation, we considered the time autocorrelation function for the radius of gyration, defined by

$$C(t) = \frac{\left\langle R_g(\tau) R_g(\tau + t) \right\rangle_\tau - \left\langle R_g(\tau) \right\rangle_\tau^2}{\left\langle R_g^2(\tau) \right\rangle_\tau - \left\langle R_g(\tau) \right\rangle_\tau^2}, \quad (22)$$

where $R_g(\tau)$ is the value of the radius of gyration at time τ, and the data are averaged over τ. Taking the AA simulation as the target, we use the reduced chi-squared statistic as a measure of the goodness of fit of the AA-RB and CG simulations,

$$\chi_{R_g}^2 = \frac{1}{N_T N_L N_t - 1} \sum_i^{N_T} \sum_j^{N_L} \sum_k^{N_t} \frac{[C_{i,j}(t_k) - C_{i,j}^{AA}(t_k)]^2}{[\sigma_{i,j}(t_k)]^2 + [\sigma_{i,j}^{AA}(t_k)]^2}, \quad (23)$$

where N_T, N_L, and N_t are the number of temperatures, polymer lengths, and time steps, respectively, $C_{i,j}^{AA}(t_k)$ and $\sigma_{i,j}^{AA}(t_k)$ are the autocorrelation functions for the radius of gyration, as defined in Equation (22), and its standard deviation at time t_k, respectively, and $C_{i,j}(t_k)$ and $\sigma_{i,j}(t_k)$ are the same quantities but for either the CG or AA-RB simulation. t_k runs in steps of 16.8 ps from 0 to 4.2 ns, and i and j index the temperatures and polymer lengths simulated. The resultant $\chi_{R_g}^2$ between CG and AA is 0.97, demonstrating
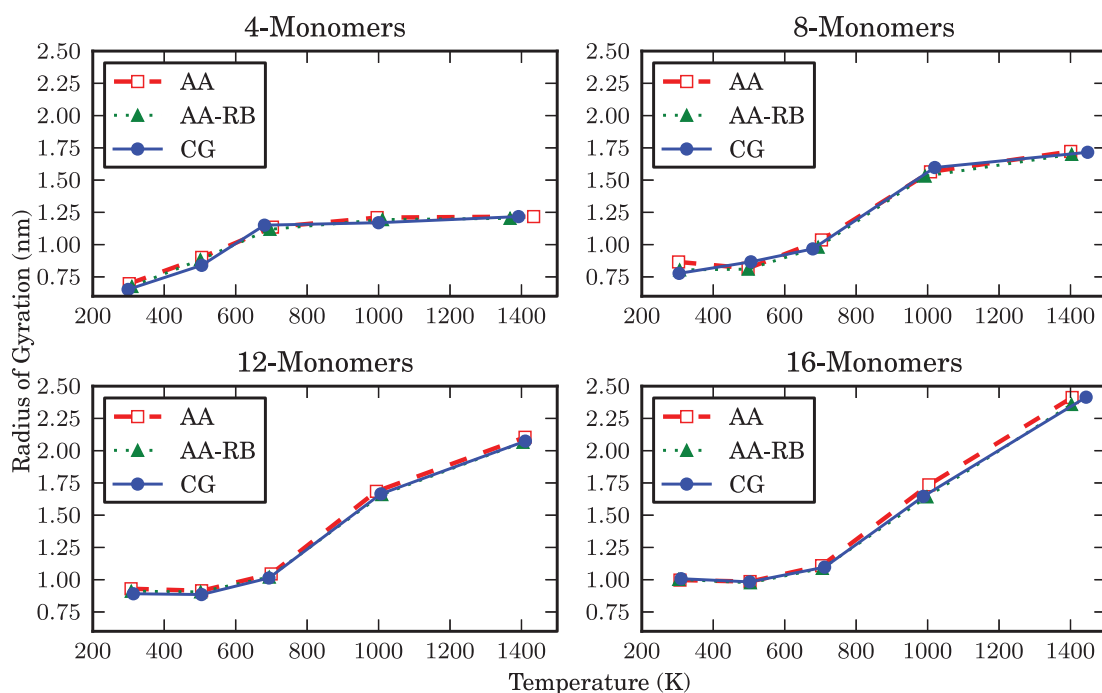

Figure 5. The radius of gyration for PEEK molecules, comprising 4, 8, 12, and 16 monomers, as a function of temperature. The CG simulations were performed with a time step of 5 fs, the AA and AA-RB simulations with a time step of 1 fs. Each data point is the average over 16 independent simulations. Straight lines have been drawn through the data to guide the eye, the AA-RB and CG data are shown as discrete symbols. Some data points lie on top of each other. Error bars are comparable to the size of the symbols.
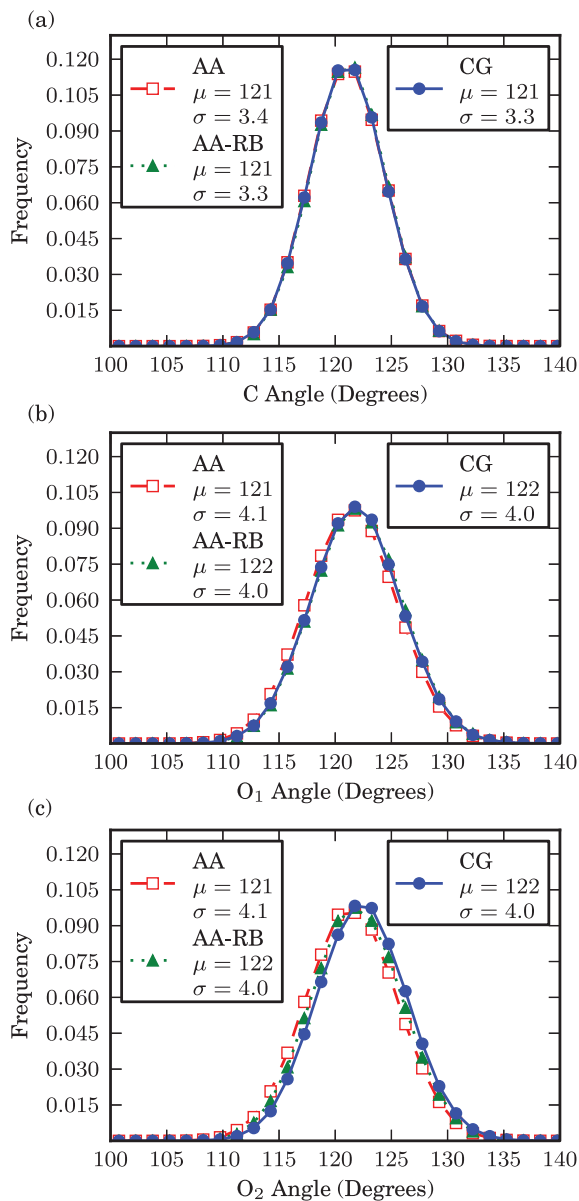
(a)

(b)

(c)

Figure 6. Normalised distributions of the hinge angles labelled C (a), O_1 (b), and O_2 (c) in Figure 2, for the 16-monomer system at 500 K, simulated in the AA (hollow squares), AA-RB (solid triangles), and CG (solid circles) representations. Some data points lie on top of each other. Lines have been added to the data as a guide to the eye. The error bars are comparable to the line width. This set of simulations has the largest discrepancy, determined by χ^2 analysis, between AA and CG representations. For many of the other systems and temperatures (not shown), the AA, AA-RB, and CG representations are indistinguishable on the scale of the plots.

excellent consistency between the dynamics of the molecules computed by the two methods. By comparison, $\chi^2_{R_g}$ between AA-RB and AA is 1.23.

In Figure 6, we show the excellent agreement between AA, AA-RB, and CG representations for the 'hinge' angles, labelled O_1, O_2, and C in Figure 2. In the ground state,
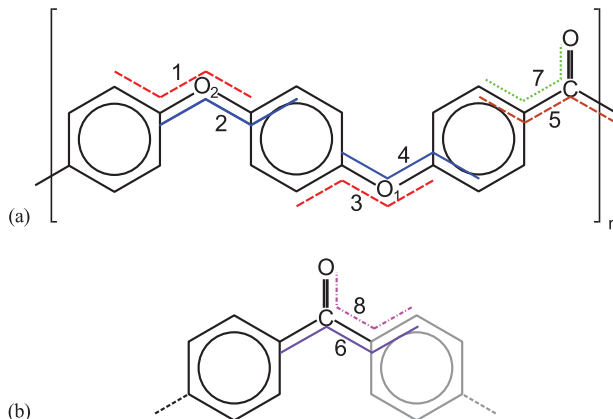

(a)

(b)

Figure 7. The dihedral angles monitored during the MD simulations of PEEK. The dihedral angles are identified by a line connecting each of the four atoms involved and are identified by a number 1–8. In (a), we see the dihedral angles which are contained within the monomer; these are 1–5 and 7. Whilst (b) shows the dihedral angles 6 and 8, which involve the first aromatic group of the next monomer (coloured grey) as it connects to the ketone. The inherent choice of atom when two atoms from the same aromatic group are used is made by mapping both possible angles into the interval $[0, 90)$, then taking the average.

these angles are $120°$. The data shown in Figure 6 are for the 16-monomer system at 500 K, which has been chosen because it displays the worst agreement between the AA and CG simulations of all the molecules and temperatures simulated; in the case of the best agreement (not shown), all data points are indistinguishable. The agreement was characterised by the reduced χ^2 value,

$$\chi^2_\theta = \frac{1}{N_a - 1} \sum_{i=1}^{N_a} \frac{\left(\theta_i^{\mathrm{CG}} - \theta_i^{\mathrm{AA}}\right)^2}{\left(\sigma_i^{\mathrm{CG}}\right)^2 + \left(\sigma_i^{\mathrm{AA}}\right)^2}, \qquad (24)$$

where θ_i^{CG} and σ_i^{CG} are the mean value and standard deviation, respectively, of the ith angle, and $N_a = 3$ is the number of angles. Smaller values of χ^2_θ indicate better agreement between data sets.

The relative orientation of the aromatic groups in PEEK is largely defined by the eight dihedral angles identified in Figure 7. There is a choice between two symmetrically equivalent atoms on each aromatic group for each dihedral angle. According to the definition of a dihedral angle, angles are measured in each of the four quadrants of a circle. The symmetry of the aromatic group results in the distribution in each quadrant being related by symmetry to the distribution in all the other quadrants. Therefore, to aid comparison between the simulations we map the measured dihedral angle $\theta \in (-180, 180]$, into the first quadrant, using the following relation:

$$\Phi = \begin{cases} |\theta| & |\theta| < 90° \\ 180° - |\theta| & |\theta| \geq 90° \end{cases}, \qquad (25)$$
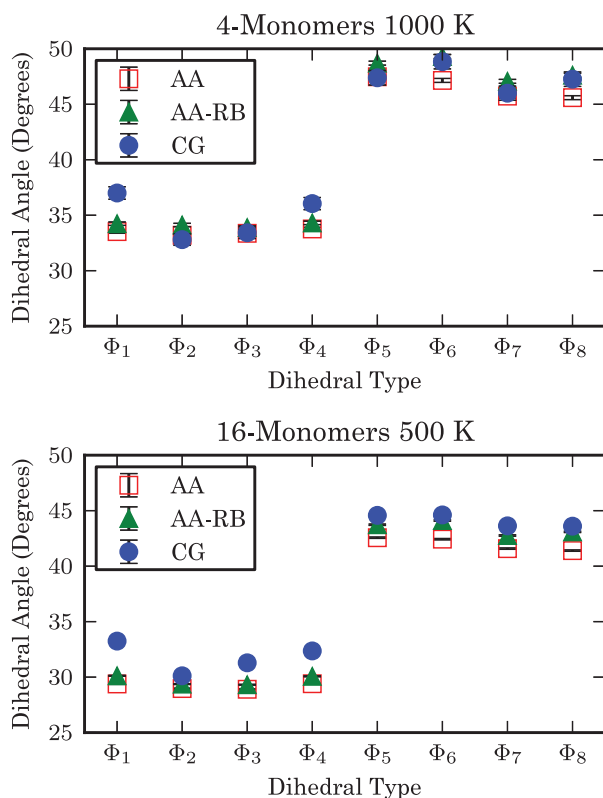
Figure 8. The mean values of the dihedral angles Φ_i labelled in Figure 7, in AA (hollow squares), AA-RB (solid triangles), and CG (solid circles) representations. The 16-monomer 500-K and 4-monomer 1000-K simulations have, respectively, the largest and the smallest reduced chi-squared values of all the configurations simulated.

where $\Phi \in [0, 90]$ is the resulting dihedral angle defining and characterising the orientation. In the case of the AA and AA-RB simulations, the aromatic group can buckle resulting in a discrepancy between the two possible dihedral angles. This discrepancy is resolved by recording the average of the two possible dihedral angles after they have been transformed by Equation (25). This reduces the effect of the buckling inherent to AA and AA-RB simulations of aromatic groups.

In Figure 8, we present the observed values for these eight dihedral angles for the 16-monomer system at 500 K and the 4-monomer system at 1000 K. These two systems were chosen because they represent, respectively, the highest and the lowest reduced chi-squared values for the dihedral angles. All angles are averaged over all dihedral angles of that type in the simulation, and over 16 independent simulations at that temperature. It can be seen that the CG dihedral angles are in excellent agreement with the AA and AA-RB simulations, and that the maximum deviation is less than $6°$, demonstrating that the orientation of the aromatic groups is preserved in the CG representation.

4. Conclusions

We have formulated and tested a coarse-graining approach to constrain planar groups of atoms along polymer backbones to move as rigid objects during MD simulations. A key feature of our method is that it preserves essential dynamical properties of each group that is CG, namely the centre of mass, the total mass, and the moment of inertia. Furthermore, the concept of 'virtual sites' is used to map forces from an AA force field on to the CG representation.

We have tested the approach by coarse-graining the aromatic groups along the backbone of the industrially important polymer PEEK over a wide range of temperatures, system sizes, and run times. The agreement with AA and AA rigid bond simulations is excellent for a number of important parameters that characterise the polymer structure. The principal limitations of our technique are those of the SHAKE algorithm: stability at large time steps and poor parallelisation.

The systematic removal of fast degrees of freedom that are irrelevant to molecular conformational changes enables the simulation to focus on the most interesting and pertinent degrees of freedom. By removing the fast out of plane vibrations associated with aromatic groups, it is possible to use an integration time step of 5 fs in the CG simulations, as compared to 1 fs in the AA simulations, for the same computational cost.

Note

1. In principle these 'atoms' could be any point masses, but to avoid confusion they will be referred to as atoms henceforth.

References

[1] K.A. Feenstra, B. Hess, and H.J.C. Berendsen, J. Comput. Chem. **20** (8), 786–798 (1999).
[2] P.L. Freddolino, F. Liu, M. Gruebele, and K. Schulten, Biophys. J. **94** (10), L75–L77 (2008).
[3] J.-P. Ryckaert, G. Ciccotti, and H.J.C. Berendsen, J. Comput. Phys. **23** (3), 327–341 (1977).
[4] H.C. Andersen, J. Comput. Phys. **52** (1), 24–34 (1983).
[5] B. Hess, H. Bekker, H.J.C. Berendsen, and J.G.E.M. Fraaije, J. Comput. Chem. **18** (12), 1463–1472 (1997).
[6] B. Hess, J. Chem. Theory Comput. **4** (1), 116–122 (2008).

[7] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, J. Chem. Theory Comput. **4** (3), 435–447 (2008).

[8] K. Anderson, R. Mukherjee, J. Critchley, J. Ziegler, and S. Lipton, Engin. Comput. **23** (1), 11–23 (2007).

[9] G. Ciccotti and J.-P. Ryckaert, Comput. Phys. Rep. **4** (6), 346–392 (1986).

[10] S. Li, D.J. Rocha, S.J. Zhou, H.S. Meyer, B. Bikson, and Y. Ding, J. Memb. Sci. **430** (0), 79–86 (2013).

[11] H. Koike, K. Kida, E.C. Santos, J. Rozwadowska, Y. Kashima, and K. Kanemasu, Tribol. Int. **49** (0), 30–38 (2012).

[12] M. Naffakh, A.M. Diez-Pascual, and M.A. Gomez-Fatou, J. Mater. Chem. **21**, 7425 (2011).

[13] G.A. Kaminski, R.A. Friesner, J. Tirado-Rives, and W.L. Jorgensen, J. Phys. Chem. B **105** (28), 6474–6487 (2001).

[14] H.J.C. Berendsen and W.F. Gunsteren, in *Molecular Liquids*, Vol. 135, *NATO ASI Series*, edited by A.J. Barnes, W.J. Orville-Thomas, and J. Yarwood (Springer, The Netherlands, 1984), pp. 475–500.

[15] H.J.C. Berendsen, D. van der Spoel, and R. van Drunen, Comput. Phys. Commun. **91** (1–3), 43–56 (1995).

[16] E. Lindahl, B. Hess, and D. van der Spoel, J. Mol. Model. **7** (8), 306–317 (2001).

[17] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, and H.J.C. Berendsen, J. Comput. Chem. **26** (16), 1701–1718 (2005).

[18] G. Bussi, D. Donadio, and M.e Parrinello, J. Chem. Phys. **126** (1), 014101 (2007).

[19] N. Michaud-Agrawal, E.J. Denning, T.B. Woolf, and O. Beckstein, J. Comput. Chem. **32** (10), 2319–2327 (2011).