Erschienen in: Soft Matter ; 5 (2009), 22. - S. 4357-4366
https://doi.org/10.1039/b912027k

Multiscale simulation of soft matter systems – from the atomistic to the coarse-grained level and back

Christine Peter and Kurt Kremer*

Many physical phenomena and properties of soft matter systems such as synthetic or biological materials are governed by interactions and processes on a wide range of length- and time-scales. Computer simulation approaches that are targeted at questions in these systems require models which cover these scales and the respective levels of resolution. Multiscale simulation methods combine and systematically link several simulation hierarchies so that they can address phenomena at multiple levels of resolution. In order to reach the mesoscopic time- and length-scales important for many material properties, methods that bridge from the atomistic (microscopic) to a coarser (mesocopic) level are developed. Here, we review coarse-grained simulation models that are linked to a higher resolution atomistic description. In particular, we focus on structure-based coarse-graining methods which are used for a variety of soft matter problems – ranging from structure-formation in amorphous polymers to biomolecular aggregation. It is shown that by coarse-grained simulation in combination with an efficient backmapping methodology one can obtain well-equilibrated long time- and large length-scale atomistic structures of polymeric melts or biomolecular aggregates which can be used for comparison to experimental data. Methodological aspects are addressed such as the question of the time-scales and dynamics in the different simulation hierarchies and an outlook to future challenges in the area of resolution exchange approaches and adaptive resolution models is presented.

Max Planck Institute for Polymer Research, Mainz, Germany. E-mail: kremer@mpip-mainz.mpg.de

1 Introduction

Material properties and behavior of soft matter systems are determined by processes and interactions on a wide range of

Christine Peter

Christine Peter studied chemistry and mathematics at the University of Freiburg, Germany. In 2003 she obtained her PhD at the ETH Zurich working on molecular dynamics simulation of biomolecular systems, and has subsequently worked at the National Institutes of Health (Bethesda, USA) and the Max Planck Institute for Polymer Research (Mainz) as a postdoctoral researcher. Currently, she is heading an Emmy Noether junior research group at the Max Planck Institute for Polymer Research working on the development of multiscale simulation models to study structure formation and self-assembly in biological and biomimetic materials.

Kurt Kremer

Kurt Kremer studied physics at the University of Cologne. In 1983 he received his PhD degree in theoretical physics from the University of Cologne on work carried out at the National Research Center KFA Jülich, where he performed computer simulations of dynamic and static properties of polymers in bulk and near surfaces. After a postdoctoral stay at the Exxon Research Center (Annandale, USA), he moved to the University of Mainz as an Assistant Professor in Theoretical Physics, after which he returned to the KFA Jülich as a senior scientific staff member. He then spent several extended periods as a visiting professor/scientist at Exxon Research and Engineering Co. (UC Santa Barbara), the University of Minnesota and the central chemical engineering department of Bayer AG (Leverkusen), looking into the applicability of current theoretical results to industrial problems. Kurt Kremer joined the Max Planck Society in 1995 as the sixth director of the Max Planck Institute for Polymer Research, heading the new theory group.

length- and time-scales. Materials are "soft" because of a characteristic low energy density, which to a very first rough approximation resembles the elastic constants. This low energy density goes in hand with locally relevant length-scales of a few Ångstroms to a few nanometres, allowing for large thermally driven fluctuations. Local, chemically specific interactions such as the specific attraction of certain units to surfaces or hydrogen bonding in aqueous environment, may be as relevant as mesoscale effects such as large-scale conformational fluctuations, hydrodynamic interactions or the formation of mesoscopic superstructures. Molecular simulation approaches to soft matter problems that are determined by such a wide range of scales demand for an equally wide range of simulation methods at various levels of resolution including a varying amount of degrees of freedom.[1,2] A variety of quantum-mechanical methods are used to address electronic/energetic properties on a high-resolution microscopic level, although these are limited to short length- and time-scales. Classical atomistic force- field methods as well as particle-based coarse-grained approaches are capable of sampling microscopic to mesoscopic scales. In particular, the latter is also capable of studying entropic effects of fluctuations, yet they still fail to cover many macroscopic properties and phenomena. For this, one needs to go beyond (purely) particle-based approaches and use, for example, the lattice Boltzmann method[3] or other mesoscopic methods to include hydrodynamic effects.

In many cases questions can be studied on a single level of resolution. A typical classical example was the question of the exponent ν for the chain extension $R^2(N) \propto N^{2\nu}$, N being the length of a polymer chain. This is a generic result, which holds independent of the underlying chemical details and one only has to distinguish good, marginal (so-called Θ) and poor solvents. However, when it comes to a quantitative understanding of complex materials, approaches with a single level of resolution do frequently not suffice since the different levels of resolution are more intimately interwoven. Here, multiscale simulation approaches are required. "Multiscale simulation" refers to methods where different simulation hierarchies are combined and linked to obtain an approach that simultaneously addresses phenomena or properties of a given system at several levels of resolution and consequently on several time- and length-scales. Multiscale simulation approaches may operate in different ways in terms of combining the individual levels of resolution: (*i*) in sequential approaches the simulation models on different scales are treated separately by simply passing information (structures, parameters, energies *etc.*) from one level of resolution to the next, (*ii*) in hybrid simulations different levels of resolution are present simultaneously, thus requiring direct interaction between them, and (*iii*) adaptive methods allow for individual molecules to adaptively switch between resolution levels on the fly – for example, depending on their spatial coordinates. In either case, the exchange of information, interaction or particles requires a high level of consistency between the individual models, which in principle have to be viewed as different physical/chemical systems that need to be adjusted to each other.

In the present short Review Article we focus on methods that have been developed to design simulation models in such a way that they are consistent with an underlying higher resolution model with the purpose to be used in a multiscale simulation framework of complex soft matter systems – synthetic as well as biological materials. More precisely, we focus on the development of coarse-grained (CG) models, which are an important intermediate level and consequently a crucial ingredient of multiscale simulation approaches.[4–9]

Simplified/generic coarse-grained models which only account for a minimal set of properties of the (macro-)molecules of interest, such as excluded volume, connectivity and a few basic types of interactions, have since long been used and are perfectly well-suited to study generic properties of soft matter systems. Since they reduce the computational complexity, they allow for much longer effective time- and length-scales than more detailed models. Good examples are the investigation of scaling properties of polymeric systems,[10] both static and dynamic, as well as the investigation of biomembranes.[11,12] For example, for the problem of polymer melt dynamics, such simulations (both molecular dynamics in continuum and Monte Carlo on lattices) have been instrumental for a better understanding of the entanglement problem.[13–15] In order to link the results of such coarse-grained simulations to real chemical systems, one needs to appropriately devise the model parameters and interaction potentials. In multiscale simulations, where one wants to switch between resolution levels or use them next to each other, one has to go beyond scaled or fitted parameters because the levels of resolution need to be linked structurally and thermodynamically consistently. This requires a very careful development of CG models to avoid unphysical effects upon changes between scales. For this reason we limit the discussion in this review to CG models which are not standing alone but which are linked to models at other (higher) levels of resolution (scale-bridging). Mostly we will focus on methodologies to develop CG models based on an atomistic (forcefield) description.

First we will discuss various methods to link the levels of resolution, these methods can be distinguished by the target (structure or thermodynamics) as well as the methodology of linking (forces, constructing effective interaction potentials). After that we will more closely investigate those methods where the coarse-grained models are developed to reproduce structural properties of an underlying higher (atomistic) description. Next to the development of the different resolution models, another important aspect is the switching between them, most notably the reinsertion of higher resolution coordinates into a CG structure. Various approaches are reviewed in the subsequent section. There we will also give a few examples of successful structures, which allow one to compute properties that can directly be compared to experiment. Then we also will address the problem of mapping the time-scales of the evolution of the systems. Finally, we will briefly touch on several additional aspects of multiscale modeling and coarse-graining, such as hybrid and adaptive resolution methods *etc.*, and give an outlook including future challenges.

2 Linking levels of resolution: energies, forces and structures

Irrespective of the method to combine the individual scales, it is an important property of a "true" multiscale simulation approach that the individual models on different levels of resolution are systematically linked. This scale-bridging requires

systematic development of the individual models such that they are thermodynamically and/or structurally consistent. As we will illustrate below, this is a challenge for multiscale modelling that needs special attention.

Many different approaches have been followed to obtain systematically linked simulation models on different levels of resolution, both from the quantum-mechanical to the classical level and from the classical all-atom level to a coarse-grained description. For the latter we discuss here a few characteristic approaches.

The derivation of interaction potentials between the coarse-grained particles may be targeted at reproducing thermodynamic properties such as energies or free energies, for example partitioning data.[16,17] This approach has proven to be particularly useful for the simulation of processes such as lipid membrane association where the said properties play a decisive role.[18] On the other hand, the energy-based coarse-graining approach does not *per se* guarantee reproduction of the structure of the system (for example of the underlying atomistic structure).[19] This may potentially cause problems and disruptions if one wishes to reinsert atomistic details into the CG structure. In contrast, structure-based methods provide CG interactions that reproduce a pre-defined target structure – often described by a set of radial distribution functions obtained from all-atom molecular simulations.[20–22] While these structure-based methods by construction reproduce local structures and thus are well-suited to reinsert atomistic coordinates, it is not *a priori* clear whether they are equally well-suited to reproduce thermodynamic properties of the system. Note that there is currently intensive research being carried out to investigate, whether – and if yes, how – it is possible to derive coarse-grained potentials that are both thermodynamically as well as structurally consistent with the underlying higher-resolution description.[23]

At first sight, a principally different methodology to construct CG interactions from an underlying atomistic simulation is the force-matching method, which has been applied to a multitude of soft matter systems, in particular biomolecular systems.[6,24] Here, the CG forcefield is determined such that the difference between the (instantaneous) CG forces and the forces in the underlying atomistic system are minimized. It can be shown that this method (in principle) determines a many-body multidimensional potential of mean force describing the CG representation of the system, thus being related to other structure-based CG methods, which usually rely on pair potentials of mean force.[25] The rather global (multibody) structural representation, however, bears the problem that the link to the underlying structure and the reproduction of local structural properties such as pair distributions may be rather weak. An exact reproduction of the underlying atomistic problem by force-matching potentially requires the introduction of higher-order interactions and forces.[26]

3 Structure-based coarse-graining – from polymers to biomolecules

The general aim in structure-based coarse-graining is to reproduce structural properties, either determined experimentally or, as in the cases we consider here, from a higher-resolution (atomistic) simulation. This means that one constructs a CG model such that it reproduces distributions of properties such as conformational degrees of freedom or (intermolecular) distances between groups of atoms which are obtained, for example, from atomistic sampling. Here, one can distinguish between several principal strategies, which also differ in the way they are linked to the microscopic system one eventually wants to understand. One possibility is to fit parameters of given potential functions such that the resulting CG simulations fit the target distributions as well as possible. Another large family of approaches uses methods to construct numerical (tabulated) potentials, based on the microscopic input, in such a way that the CG model exactly reproduces the target structure. The latter group of methods relies in principle on Boltzmann inversion of the distributions – in combination with some iterative procedure to remove multibody effects, as will be explained in more detail below.

Below we will, in addition to references to the literature, mostly resort to work of the Mainz group, where we illustrate the increasing complexity of CG models and scale-bridging approaches: (*i*) BPA-PC,[4,27] and (*ii*) polystyrene[28–31] are typical examples for amorphous polymeric systems, polymers being among the first systems for which this development of CG models for multiscale simulation purposes was ventured. (*iii*) With the low molecular weight liquid-crystalline compound 8AB8 we show how the recipes from the polymer world can be extended to systems, where chemically specific nonbonded interactions play an increasingly important role for structure formation.[22] (*iv*) With a dipeptide (diphenylalanine) in aqueous environment we provide an outlook to show how the CG methodology can be extended to biomolecular systems where the complexity compared to homogeneous isotropic polymer melts is substantially increased.[32,33] Fig. 1 shows the chemical structure of the named compounds. We will here explicitly concentrate on bulk properties. It should be noted though that for many scientific as well as technological questions, the interaction of a (macro-)molecular matrix with an (in)organic surface is of central relevance. Such systems have been studied using scale-bridging methods as well, both the local interaction of relatively small biomolecules with a metal surface[34,35] as well as the behavior of differently terminated

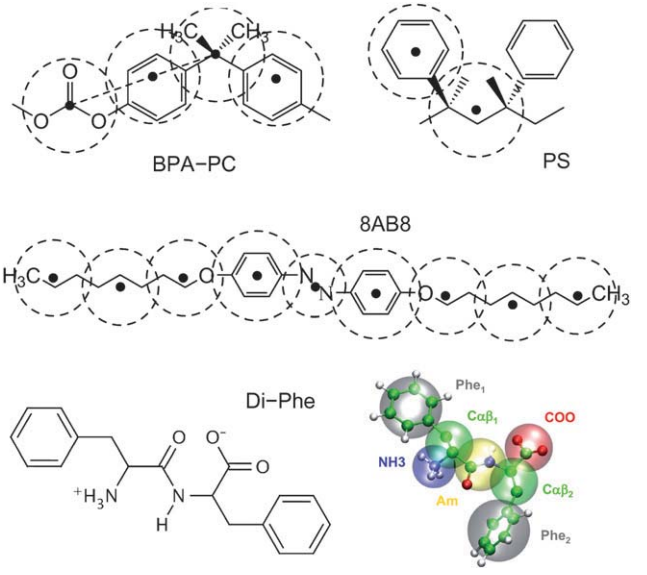

Fig. 1 Chemical structure and mapping schemes of the discussed CG examples: BPA-PC, polystyrene (PS), 8AB8, diphenylalanine.

polycarbonates close to a nickel surface, which is important in the context of optical discs.[36,37]

In many CG approaches the total set of CG interaction functions is separated into bonded/covalent and nonbonded potentials which are developed separately. This approach relies on the assumption that the total potential energy U^{CG} can be separated into the respective contributions

$$U^{CG} = \sum U_B^{CG} + \sum U_{NB}^{CG} \quad (1)$$

Following this separation ansatz, we will first discuss the derivation of bonded (covalent, intramolecular) interaction potentials and possible implications of interdependent/correlated degrees of freedom. This separation ansatz is also important in the context of transferability of the CG model. Assuming such a strict separation means that the intramolecular bonded/covalent CG interactions should be independent of the special scientific problem, *i.e.* of the surroundings of the molecules.

3.1 Bonded/covalent interaction potentials

Bonded interactions are derived such that the conformational statistics of the molecules is represented correctly in the CG model. These conformational distributions P^{CG} are usually characterized by specific bond lengths r, angles θ, and torsions ϕ between any pair, triplet or quartet of CG beads respectively, *i.e.* $P^{CG}(r,\theta,\phi,T)$. The target distributions are determined by sampling the conformational degrees of freedom at atomistic resolution (the different sampling methods will be discussed in more detail below). If one assumes that the different CG internal degrees of freedom are uncorrelated, $P^{CG}(r,\theta,\phi,T)$ factorizes into independent probability distributions of bond, angle and torsional degrees of freedom

$$P^{CG}(r,\theta,\phi,T) = P^{CG}(r,T)\, P^{CG}(\theta,T)\, P^{CG}(\phi,T) \quad (2)$$

This assumption has to be carefully checked (it is not uncommon that coarse-grained degrees of freedom are correlated, for example that certain combinations of CG bonds, angles and torsions are "forbidden" in the distributions obtained from the atomistic sampling), and is an important test of the suitability of a mapping scheme, *i.e.* the relation between atomistic and CG centers. A mapping scheme that requires complex multi-parameter potentials is computationally rather inefficient. For a detailed discussion of how this can be achieved for the rather complex problem of different stereoregular subunits of polystyrene, we refer the reader to a recent study by Fritz *et al.*[31] The individual probability distributions $P^{CG}(r,T)$, $P^{CG}(\theta,T)$, and $P^{CG}(\phi,T)$ are then Boltzmann-inverted to obtain the corresponding potentials:

$$U^{CG}(r,T) = -k_B T \ln(P^{CG}(r,T)/r^2) + C_r \quad (3)$$

$$U^{CG}(\theta,T) = -k_B T \ln(P^{CG}(\theta,T)/\sin(\theta)) + C_\theta \quad (4)$$

$$U^{CG}(\phi,T) = -k_B T \ln P^{CG}(\phi,T) + C_\phi \quad (5)$$

with C_r, C_θ, and C_ϕ being irrelevant constants used to set the minima of the respective potentials to zero. Note that these potentials are in fact potentials of mean force (and therefore free energies), and consequently are temperature dependent (not only due to the prefactor $k_B T$). This means they can strictly be only applied at the temperature (state point) they were derived at, requiring a reparametrization at each temperature. In practice, one needs to test the width of this applicability range for each CG model individually. Experience shows that typical temperature ranges are of the order of ±10–20% (if no phase transition is within that range). Technically, the Boltzmann inversions (eqns 3–5) and the subsequent determination of the derivatives can be carried out numerically, resulting in tabulated potentials and forces. Another option is to determine analytical potentials that reproduce the probability distributions $P^{CG}(r,T)$, $P^{CG}(\theta,T)$, and $P^{CG}(\phi,T)$, for example by fitting the (multipeaked) bond and angle distributions by a series of Gaussian functions which can then be inverted analytically, resulting in smooth potentials and forces.[22,38]

If one aims at maintaining a systematic separation of bonded and nonbonded degrees of freedom, the conformational sampling at atomistic resolution needs to be performed using isolated molecules (for example using MD simulations with a stochastic thermostat). This has been successfully applied to several amorphous polymeric systems, where bond, angle and torsional distributions are determined from the conformations of a single polymer chain *in vacuo*.[22,28,29] When these reference distributions are generated, the inclusion of nonbonded interactions has to be taken with care to avoid "double counting" of long range intra-chain interactions, as explained in more detail in ref. 9.

It is, however, also possible to determine the CG internal degrees of freedom based on distributions obtained from an atomistic simulation of the polymer melt.[21] In the latter case one obtains potentials for bonded and nonbonded interactions simultaneously based on the same melt (through iteration as described below). Consequently all interaction functions are interdependent, *i.e.* there is no clear separation between covalent and nonbonded interaction potentials. This makes (for example) the detection of correlations between different bonded degrees of freedom – which are very important for the correct representation of local structures – practically impossible. In addition, the separation of bonded (intramolecular) and nonbonded degrees of freedom is necessary if one wants to reuse nonbonded potentials for certain chemical units (fragments of the target molecule; see below).

While the above clear separation of bonded and nonbonded interactions is desirable from a statistical mechanical point of view, the derivation of meaningful bonded potentials from a single molecule *in vacuo* can only be successful if the conformational sampling of the molecule *in vacuo* and in the bulk (or solution) phase do not differ substantially. In biomolecular systems, due to the peculiar nature of aqueous solutions (*i.e.* the presence of hydrogen bonds), this assumption gets problematic, as is illustrated by the dipeptide diphenylalanine.[32] For this dipeptide, bonded CG potentials were determined by Boltzmann inversion of the respective distributions obtained from conformational sampling of a single peptide in aqueous solution. Note that even though bonded and nonbonded interactions are not as rigorously separated as in the polymer examples (due to the conformational sampling in solution instead of vacuum environment), covalent and nonbonded interactions were

nevertheless separately and sequentially determined. This means that there is still the possibility to analyze and incorporate correlations into the conformational sampling and to reuse nonbonded interaction functions determined from molecule fragments. The resulting CG model of the dipeptide (after adding also nonbonded potentials) turned out to reproduce very well the conformational equilibrium of the atomistic peptide. Correlations between bonds, angles, and torsions as well as overall conformational properties such as the end-to-end distance were also reproduced by the CG model.[32,33] It should be noted, though, that this cannot be taken for granted and needs to be verified anew for different biomolecules. This model validation becomes even more important if further intramolecular multibody interactions (for example improper torsion angle potentials to preserve stereocenters) are required to reproduce the conformations of the molecule. In that case, several interaction potentials may affect the same sets of CG beads; they are therefore not independent and the separability and additivity assumed in eqn 2 is no longer warranted. For more details on this, see ref. 32.

3.2 Nonbonded interaction potentials

Nonbonded interactions can be introduced in a variety of ways, depending on the system and the question one is studying. For amorphous polymers, where the density is known from experiment or atomistic simulations, it is in many cases sufficient just to introduce an appropriate excluded volume for the CG beads. This approach has been used successfully for studies on polycarbonate and polystyrene[4,27,28] and is in line with very early ideas, where specific polymers have even been studied by lattice models.[39–41] In all other cases nonbonded interaction potentials between coarse-grained beads are derived based on the structure of isotropic liquids of small molecules (in the case of more complex molecules such as the liquid-crystalline compound 8AB8, fragments of the target molecule are used). In this case, radial distribution functions of the atomistically simulated liquids are used as targets for the parameterization process. As mentioned, there are two principal options to determine and optimize CG potentials: the first option is to adjust the parameters of analytical potentials such as Lennard-Jones to closely reproduce the structure of the atomistic melt/liquid (for examples and more details see refs. 22 and 29). The second option is to use numerically derived tabulated potentials which are designed such that the CG melt exactly reproduces the atomistic melt structure.

In the second scheme, the inverse Monte Carlo or the iterative Boltzmann inversion method[20,42] can be used to numerically generate a tabulated potential that precisely reproduces a given radial distribution function $g(r)$. These methods rely on an initial guess for a nonbonded potential $U^{\mathrm{CG}}_{\mathrm{NB},0}$. Often the Boltzmann inverse of the target $g(r)$, *i.e.* the potential of mean force,

$$U^{\mathrm{CG}}_{\mathrm{NB},0} = -k_{\mathrm{B}}T\ln g(r) \tag{6}$$

is used, with which one then performs a coarse-grained simulation of the liquid. The resulting structure of this first step will not match the target structure as the potential of mean force is – due to multibody interactions – only in the limit of very high dilution a good estimate for the potential energy. The iterative Boltzmann method uses the following iteration scheme

$$U^{\mathrm{CG}}_{\mathrm{NB},i+1} = U^{\mathrm{CG}}_{\mathrm{NB},i} + k_{\mathrm{B}}T\ \ln\left(\frac{g_i(r)}{g(r)}\right) \tag{7}$$

with which the original guess can be self-consistently refined until the desired structure is obtained.

At this stage it should be mentioned that it can be shown[43] that the solution to the problem of finding a pair potential which exactly reproduces a given radial distribution function is unique. In practice, however, there may exist several different pair potentials (in particular in the long-range attractive tail) that reproduce a given structure function with a hardly noticeable error. This property opens up the chance to impose additional constraints to the Boltzmann inversion procedure in such a way that other thermodynamic quantities are reproduced, for example the pressure[42,44–46] or the compressibility of a liquid,[46] without disrupting the local structure. This will surely be subject of future investigations, since it touches on the question to what extent structure-based CG potentials are also capable of reproducing important thermodynamic properties of a system.[23,45] This is of particular importance if one wants to use such structure-based potentials for applications beyond the simulation of polymeric melt structures, for example biological assembly processes where partitioning properties and aspects of solubilities *etc.* play a decisive role.

For complex molecules with a large number of different CG beads (for example biological macromolecules), or in the case of liquid-crystalline molecules with anisotropic structures, the procedure to determine nonbonded interaction functions is slightly more involved. In these cases it is advantageous to split the target molecule into fragments so that the nonbonded interactions between different bead types can be determined based on the structure of *isotropic* liquids or mixtures/solutions of these fragment molecules. This fragment-based approach has been successfully applied to the liquid-crystalline compound 8AB8[22] (see also Fig. 1). A potential error that could be introduced in the parametrization based on molecule/chain fragments is that different conformations or relative orientations between molecules might contribute differently to the structure of the fragment liquids than to the target liquid or melt.[47] Consequently these conformations could be misrepresented in the CG potentials. One example is the relative weights of parallel and perpendicular orientations between anisotropic molecules such as phenyl rings, which might differ in liquid benzene compared to molecules where the rings are embedded into a longer chain. This and other aspects of transferability of CG potentials, for example the transferability between different compositions of liquid/liquid mixtures, have to be carefully tested, always keeping in mind that in principle all CG potentials are state-dependent. Nevertheless, the procedure to derive CG potentials from chain fragments and low molecular weight liquids opens up the possibility to reuse certain CG potentials for re-occurring building blocks (such as alkyl or phenyl groups). Fig. 2 illustrates the result of the structure-based coarse-graining approach for the peptide–water interactions in diphenylalanine. For each CG peptide bead the interaction with CG water had been determined by iterative Boltzmann inversion – based on atomistic sampling

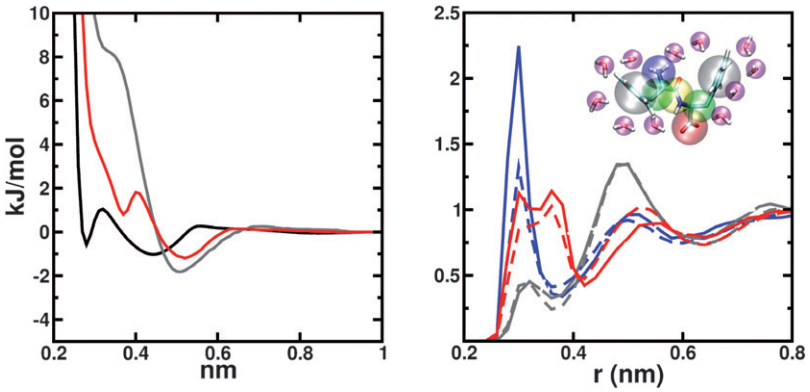

Fig. 2 Structure-based coarse-graining of the peptide–water interactions in diphenylalanine by iterative Boltzmann inversion based on atomistic simulations of peptide fragments (after ref. 33). Left panel: CG nonbonded potentials from iterative Boltzmann inversion for water–water (black), Phe–water (grey), and COO–water (red) interactions. Right panel: Comparison of the solvation structure around diphenylalanine in CG and atomistic simulations using radial distribution functions between various peptide and water beads in atomistic (solid lines) and CG simulations (dashed lines). Blue lines refer to the N-terminal, red to the C-terminal and grey to the phenyl bead type. Inset: Snapshot of the CG and the back-mapped peptide–water system (only first solvation shell).

of the radial distribution function of a corresponding fragment molecule with water. The resulting potentials for a few CG beads are shown in the left panel (note that by construction the resulting CG radial distribution functions exactly match the atomistic reference distributions). The right panel of Fig. 2 shows that the solvation structure around the diphenylalanine peptide in the CG simulation (after combining the independently determined interaction potentials) reproduces well the corresponding structure in the atomistic simulation.

The previous sections have been mainly concerned with homogeneous bulk systems for which a well-equilibrated reference radial distribution function (or its Boltzmann inverse, the potential of mean force) can be easily obtained, and where the iterative procedure to eliminate multibody effects rather straightforwardly works and easily converges. If one wants to extend the methodology to biomolecular systems, one has to reproduce structures/potentials of mean force at rather low concentrations in aqueous environment. For dilute solutions, the iterative Boltzmann inversion method is not a particularly useful approach, since the solute–solute radial distribution functions (solute is here referring to the low-concentration component) converges very slowly. In this case, a direct determination of the potential of mean force (PMF) between two solute molecules in atomistic simulations using free-energy methods such as umbrella sampling or constraint dynamics is more useful. This PMF then needs to be reproduced by the CG system. Analogously to the iteration steps in the iterative Boltzmann inversion method, a procedure has been developed that constructs a pair potential that reproduces a potential of mean force in dilute solution. Instead of eliminating multi-body effects of surrounding molecules iteratively, one subtracts out the effects of the solvent molecules, as described in detail in ref. 33.With this methodology, a CG model for diphenylalanine has been developed in such a way that the coarse-grained level maintains explicit solvent degrees of freedom, the CG water model preserves the solvation structure around the peptides and the peptide–peptide interaction in water is represented correctly, resulting in the correct thermodynamic association behavior.

4 Backmapping

Various approaches have been employed to reinsert atomistic details into a CG structure or simulation trajectory. It should be noted that this "backmapping" or inverse mapping problem has in general no unique solution, since every CG structure corresponds to an ensemble of atomistic microstates. Thus, the task is to find a representative structure from this ensemble which displays the correct statistical weight of those degrees of freedom not resolved in the CG description, *e.g.* ring flips *etc.* For coarse-grained polymeric melts it is possible to obtain backmapped atomistic structures by taking rigid all-atom chain fragments – corresponding to a single or a small set of CG beads – from a correctly sampled distribution of all-atom chain structures. These fragments were fitted onto the CG structure, and the resulting all-atom structure was relaxed by energy minimization and a short equilibration.[4,28,48–50] This procedure is necessarily limited to systems where the overall structural relaxation and diffusion of molecules is slow compared to the local equilibration of the atomistic coordinates after fitting, for example polymeric melt structures. The case of more flexible low-molecular weight molecules requires a slightly different strategy in order to avoid the atomistic structure from drifting/diffusing too far from the CG reference. Several similar approaches have been developed for this case.[22,32,33,51] The general strategy is to insert an initial set of atomistic coordinates into the CG structure (either using presampled fragments or random coordinates) in such a way that these atomistic coordinates satisfy the "mapping condition", *i.e.* the atomistic coordinates have to yield back the CG structure if one would apply the mapping scheme. The resulting atomistic structure then needs to be relaxed and equilibrated. This can be done, for example, by a short molecular dynamics simulation with an additional restraining potential of groups of atoms to their corresponding CG mapping points. This procedure generates a well-equilibrated ensemble of atomistic microstates that correspond to the coarse-grained structure. Fig. 3 shows the result of the backmapping procedure for the systems that are discussed in more detail in this article: a BPA-PC and a polystyrene chain, a snapshot of liquid-crystalline 8AB8, a snapshot

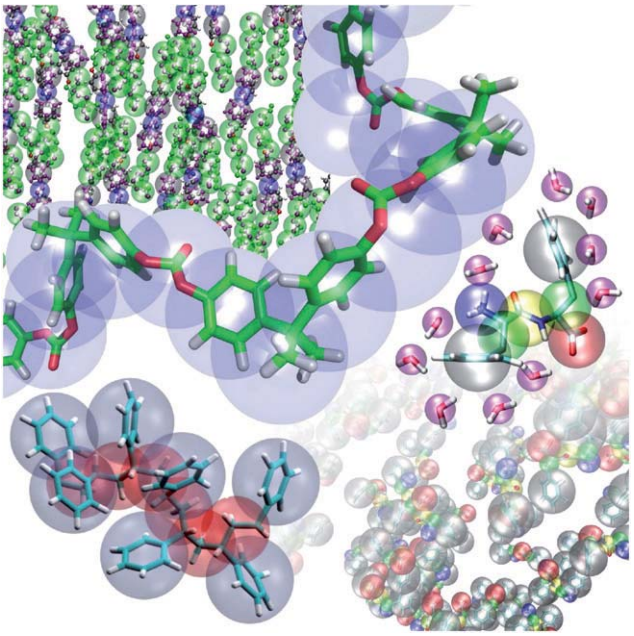

Fig. 3 Coarse-grained superimposed with backmapped atomistic structures of a BPA-PC chain and a polystyrene oligomer, of liquid-crystalline 8AB8 (upper left corner), of the aggregates formed by diphenylalanine (lower right corner) and a single diphenylalanine molecule with its water shell.

of the aggregates formed by diphenylalanine and a single diphenylalanine molecule with its water shell.

The combination of CG simulations with an efficient backmapping methodology is a powerful tool to efficiently simulate long time-scale and large length-scale soft matter processes and in the end to obtain well-equilibrated atomistic structures and trajectories. These resulting structures can be directly compared to experimental data. For example, the relevant timescale of many NMR experiments requires simulations beyond what is possible with atomistic models. Nevertheless, atomistic coordinates are necessary to compare with experimental results, an important example for the use of backmapped CG trajectories.[48,52]

In a slightly different manner one can also utilize inverse-mapped structures in further computation. For example, in order to obtain data for solubilities or permeabilities of small molecules in polymeric systems, one can combine coarse-grained simulations to obtain well-equilibrated structures of the polymeric melt with atomistic free energy calculations based on the inverse-mapped trajectories.[53–55] In a rather early study of phenol in BPA-PC, phenol was introduced into remapped polycarbonate melts at a variety of temperatures. There it could be shown, how the diffusion of the phenols coupled to the local fluctuations of the polymeric matrix.[56]

Another promising application of the combination of coarse-grained simulations with a backmapping procedure is the possibility to validate the underlying atomistic forcefield – on time- and length-scales not accessible to atomistic simulations alone due to sampling problems.

5 Dynamics: coarse-grained *versus* atomistic

Within CG models length-scales are usually well-defined through the construction of the coarse-graining itself. However, for the question of dynamics this is not the case at all. We illustrate this here for the example of polymers. From polymer simulations of both simple continuum as well as lattice models, it is known that such simulations reproduce the essential generic features of polymer dynamics; that is, the crossover from the Rouse to the entangled reptation regime, qualitatively and to a certain extent quantitatively.[15] The simple polymer models are, in view of the present discussion, just another set of different polymers. They behave, because of their connectivity and excluded volume constraints *etc.*, just as any other polymer, irrespective of whether the repeat units are made of carbons, hydrogens *etc.*, or whether they are sites on a lattice or simple Lennard-Jones beads. Properly scaled, they all follow the same rules. For short chains the longest relaxation time (τ_R) scales as the square of the chain length and for long chains in a polymer melt we observe reptation behavior. Here, however, we would like to go beyond this level of comparison, and show that a proper link between the atomistic representation of a system and the corresponding structurally coarse-grained system can provide absolute dynamic information without the need to resort to generic scaling laws. Actually, eventually one should recover them as well. Thus our aim is predictive quantitative modeling of diffusion, viscosity, rates, and correlation times, *etc.* of dynamic events. This automatically generates the question of the minimal time- and length-scales CG simulations apply to. First it is important to realize that the coarse-grained models are from a simulation point of view independent models with their own intrinsic dynamics. In the case of the previously discussed polystyrene and BPA-PC simulations, one can deduce a typical simulation time scale, as is traditionally done in MD simulations. Taking the strength of the interaction parameter in the nonbonded excluded volume interaction ε_{CG} (measured in units of the temperature), the average mass m_{CG} of the CG beads and the known length-scales σ_{CG}, one can determine the intrinsic time scale of the CG simulation from $1\tau = 1(m_{CG}\sigma_{CG}^2/\varepsilon)^{1/2}$. This results for instance in $1\tau = 1.7$ ps for BPA-PC at 570 K and $1\tau = 1$ ps at 463 K for atactic polystyrene respectively.[30,57] Note that the resulting time scale τ depends on the choice of masses, if the CG model contains different beads. Thus there is some arbitrariness in the value of τ if one wants to use physical units. In most cases we hence prefer to only speak of "τ". While this is the natural time scale of the CG model, this does not at all have to be the time scale of the underlying atomistic model. The CG interaction potentials are much smoother, barriers are lower *etc.*, resulting in a significantly accelerated dynamics. Beyond the reduction of the number of degrees of freedom, this is the main reason for the speed-up due to coarse-graining.

On the other hand, on length-scales above the typical scale of the coarse-graining, we expect qualitatively the same behavior for the CG chains as for the atomistic chains, certainly on scales where generic properties dominate. This offers a direct way of deducing the time-scaling between the CG model and the underlying atomistic model by matching the curves of the mean-square displacements of the beads or the center of mass of the whole chain. Since the lengths are fixed by the mapping procedure itself, the mean-square displacements can be matched just by shifting the time-scales. It is, however, important that the curves of the atomistic and CG mean-square displacements not only meet in a point, but rather coincide from a characteristic

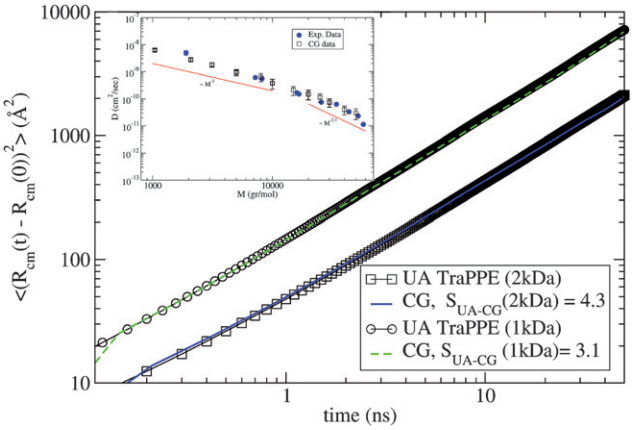

Fig. 4 Mean-square displacements of centers of mass of coarse-grained and atomistic (united-atom) polystyrene (PS) simulations as a function of time for two different chain lengths. By shifting the CG data along the time axis the time scaling can be obtained. To obtain the real time scaling, in another step the united atom (UA; here aliphatic hydrogens are lumped into carbon superatoms) and the atomistic (AA) simulations have to be compared in the same way. This is necessary since atomistic simulations for PS are extremely time-consuming. Because of this, two different atom-based methods, AA and UA respectively, are needed to obtain an initial scale factor from distances, which are too small for CG simulations. The inset shows the diffusion constants for PS obtained from CG simulations based on the resulting AA-CG time-scaling in comparison to an experiment for a molecular weight of up to $M = 50$ kDa (after refs. 30 and 58). Note that there is no adjustable parameter in the simulation data.

point on onwards, as will be discussed below. Fig. 4 shows a typical result for polystyrene.[58] This procedure leads for the two examples mentioned above to the time scaling of $1\tau = 400$ ps for PS at 463 K and 30 ps for BPA-PC at 570 K respectively. Of course, as this example already illustrates, different levels of coarse-graining and different CG models lead to different time-scales. Also it should be noted that the time-scaling factor is chain-length-dependent, since the melt density varies as function of chain length for polymers below the entanglement molecular weight (see Fig. 4).

The data upon which Fig. 4 is based reveal another important information. Looking at the displacements of the individual beads, rather than the center of mass of the chains, one finds that the motion characteristics (down to the characteristic scale of our coarse-graining) agree (after the appropriate time scaling) qualitatively and quantitatively with the atomistic simulations. Thus the CG runs can be used to obtain realistic atomistic trajectories. In ref. 48 it was shown for BPA-PC that atomistic dynamic structure factors can be obtained from remapped CG runs in perfect match with short-time atomistic data, and by this extend the dynamic information significantly. Recently, we have shown for polystyrene that this also quantitatively predicts correct bond orientation correlation times as obtained from NMR.[58]

Methods like this not only can be used to study the dynamics of homopolymer melts, but also the dynamics of additives in such melts.[59,60] A recent study of ethylbenzene in polystyrene without any adjustable parameter perfectly reproduces the Vogel–Fulcher behavior of the ethylbenzene diffusion as a function of temperature and allows quantitative predictions into regions that are experimentally extremely difficult to access. However, such an analysis is somewhat more involved, as the time scaling for the additive and for the chain are usually not the same. This requires special attention, and a detailed discussion is beyond the scope of this short review.

6 Conclusions and outlook

In this review we have summarized a number of methods to devise coarse-grained models. The general challenge that lies in coarse-graining and generally in reducing the number of degrees of freedom of a computational model is to incorporate the (average) effect of the eliminated degrees of freedom into the lower-resolution model. We have summarized different developments which are currently being pursued with the aim to design coarse-grained models that are both thermodynamically and structurally consistent with an underlying atomistic simulation model. This consistency is of particular importance for coarse-grained models that are to be used in a multiscale simulation framework where different simulation hierarchies are combined and linked to obtain an approach that simultaneously addresses phenomena or properties of a given system at several levels of resolution and consequently on several time- and length-scales. Such multiscale approaches are of great importance in the investigation of complex soft matter systems such as biological and synthetic materials, where phenomena on a wide range of scales "team up" to determine the overall (material) properties.

The development of multiscale simulation methods is one of the major methodological efforts in computational chemistry and physics. These scale-bridging approaches can operate on varying levels of "interaction" between the individual scales: the examples shown in the present article mainly combine models on different scales by treating them separately and sequentially, *i.e.* they pass information (in the present cases structures) between the levels of resolution. Resolution exchange methods use the exchange (of the whole system) between simulations at different levels of resolution during the course of the simulation as a means to enhance sampling.[61–63] In the case of hybrid simulations, different levels of resolution are present simultaneously in one system. This is more complex than the sequential approach, since interactions between entities at different levels of resolutions have to be devised. Hybrid approaches are widely used in the field of mixed quantum-mechanical/classical simulations. The problem if one wants to use hybrid approaches in mixed classical atomistic/coarse-grained simulations is that diffusion plays a greater role here, so that one faces the question how to treat particles that leave one region of a certain resolution and enter another. A treatment of such problems that is consistent with statistical mechanics has been addressed in methods that allow for individual molecules to adaptively switch between resolution levels on the fly.[44,64–66] Ongoing development of these and other multiscale simulation methods and their application to problems related to structure formation, self-assembly, and surface interactions in synthetic and biological systems pose a multitude of challenges in computational chemistry and physics.

Acknowledgements

We thank all present and past members of the multiscale modeling group at the Max Planck Institute for Polymer Research for fruitful collaborations and many stimulating discussions, in particular Luigi Delle Site, Vagelis Harmandaris, Nico van der Vegt and Alessandra Villa. We also thank Nico van der Vegt for carefully reading the manuscript and many helpful suggestions. We would like to acknowledge various teams developing multiscale modeling software packages, namely the Espresso and Espresso++ team,[67] the Gromacs developers[68] (most notably Berk Hess for implementing numerous CG methods), and the VOTCA[26] team for putting together a powerful coarse-graining suite. CP acknowledges financial support by the German Science Foundation within the Emmy Noether Programme (grant PE 1625/1-1).

References

1 H. J. C. Berendsen, *Simulating the Physical World*, Cambridge University Press, Cambridge, UK, 2007.
2 *Coarse-Graining of Condensed Phase and Biomolecular Systems*, ed. G. A. Voth, Chapman and Hall/CRC Press, Taylor and Francis Group, 2008.
3 B. Dunweg and A. J. C. Ladd, Advanced Computer Simulation Approaches for Soft Matter Sciences III (Advances in Polymer Science, vol. 221), 2009, pp. 89–166.
4 W. Tschöp, K. Kremer, J. Batoulis, T. Burger and O. Hahn, *Acta Polym.*, 1998, **49**, 61–74.
5 W. Tschöp, K. Kremer, O. Hahn, J. Batoulis and T. Burger, *Acta Polym.*, 1998, **49**, 75–79.
6 G. S. Ayton, W. G. Noid and G. A. Voth, *Curr. Opin. Struct. Biol.*, 2007, **17**, 192–198.
7 P. L. Freddolino, A. Arkhipov, A. Y. Shih, Y. Yin, Z. Chen and K. Schulten, *Coarse-Graining of Condensed Phase and Biomolecular Systems*, Chapman and Hall/CRC Press, Taylor and Francis Group, 2008.
8 S. J. Marrink, M. Fuhrmans, H. J. Risselada and X. Periole, *Coarse-Graining of Condensed Phase and Biomolecular Systems*, Chapman and Hall/CRC Press, Taylor and Francis Group, 2008.
9 N. F. A. van der Vegt, C. Peter and K. Kremer, *Coarse-Graining of Condensed Phase and Biomolecular Systems*, Chapman and Hall/CRC Press, Taylor and Francis Group, 2008.
10 K. Kremer and G. S. Grest, *J. Chem. Phys.*, 1990, **92**, 5057–5086.
11 M. Muller, K. Katsov and M. Schick, *Phys. Rep.*, 2006, **434**, 113–176.
12 B. J. Reynwar, G. Illya, V. A. Harmandaris, M. M. Müller, K. Kremer and M. Deserno, *Nature*, 2007, **447**, 461–464.
13 T. C. B. McLeish, *Adv. Phys.*, 2002, **51**, 1379.
14 R. Everaers, S. K. Sukumaran, G. S. Grest, C. Svaneborg, A. Sivasubramanian and K. Kremer, *Science*, 2004, **303**, 823–826.
15 K. Kremer, *Simulations in Condensed Matter: From Materials to Chemical Biology*, Springer, 2006, vol. 704.
16 S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman and A. H. de Vries, *J. Phys. Chem. B*, 2007, **111**, 7812–7824.
17 L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman and S. J. Marrink, *J. Chem. Theory Comput.*, 2008, **4**, 819–834.
18 S. J. Marrink, A. H. de Vries and A. E. Mark, *J. Phys. Chem. B*, 2004, **108**, 750–760.
19 R. Baron, D. Trzesniak, A. H. de Vries, A. Elsener, S. J. Marrink and W. F. van Gunsteren, *ChemPhysChem*, 2007, **8**, 452–461.
20 A. P. Lyubartsev and A. Laaksonen, *Phys. Rev. E*, 1995, **52**, 3730–3737.
21 F. Müller-Plathe, *ChemPhysChem*, 2002, **3**, 754–769.
22 C. Peter, L. Delle Site and K. Kremer, *Soft Matter*, 2008, **4**, 859–869.
23 M. E. Johnson, T. Head-Gordon and A. A. Louis, *J. Chem. Phys.*, 2007, **126**, 144509.
24 S. Izvekov and G. A. Voth, *J. Phys. Chem. B*, 2005, **109**, 2469–2473.
25 W. Noid, J. Chu, G. Ayton and G. Voth, *J. Phys. Chem. B*, 2007, **111**, 4116–4127.
26 V. Ruehle, C. Junghans, A. Lukyanov, K. Kremer and D. Andrienko, *J. Chem. Theory Comput.*, 2009, submitted.
27 C. F. Abrams and K. Kremer, *Macromolecules*, 2003, **36**, 260–267.
28 V. A. Harmandaris, N. P. Adhikari, N. F. A. van der Vegt and K. Kremer, *Macromolecules*, 2006, **39**, 6708–6719.
29 V. A. Harmandaris, D. Reith, N. F. A. van der Vegt and K. Kremer, *Macromol. Chem. Phys.*, 2007, **208**, 2109–2120.
30 V. A. Harmandaris and K. Kremer, *Macromolecules*, 2009, **42**, 791–802.
31 D. Fritz, V. A. Harmandaris, K. Kremer and N. F. A. van der Vegt, *Macromolecules*, submitted.
32 A. Villa, C. Peter and N. F. A. van der Vegt, *Phys. Chem. Chem. Phys.*, 2009, **11**, 2077.
33 A. Villa, N. F. A. van der Vegt and C. Peter, *Phys. Chem. Chem. Phys.*, 2009, **11**, 2068.
34 P. Schravendijk, L. M. Ghiringhelli, L. Delle Site and N. F. A. van der Vegt, *J. Phys. Chem. C*, 2007, **111**, 2631–2642.
35 L. M. Ghiringhelli, B. Hess, N. F. A. Van der Vegt and L. Delle Site, *J. Am. Chem. Soc.*, 2008, **130**, 13460–13464.
36 L. Delle Site, S. Leon and K. Kremer, *J. Am. Chem. Soc.*, 2004, **126**, 2944–2955.
37 D. Andrienko, S. Leon, L. Delle Site and K. Kremer, *Macromolecules*, 2005, **38**, 5810–5816.
38 G. Milano, S. Goudeau and F. Muller-Plathe, *J. Polym. Sci., Part B: Polym. Phys.*, 2005, **43**, 871–885.
39 W. Paul, K. Binder, K. Kremer and D. W. Heermann, *Macromolecules*, 1991, **24**, 6332–6334.
40 P. Doruker and W. L. Mattice, *Macromolecules*, 1997, **30**, 5520–5526.
41 V. Tries, W. Paul, J. Baschnagel and K. Binder, *J. Chem. Phys.*, 1997, **106**, 738–748.
42 D. Reith, M. Putz and F. Muller-Plathe, *J. Comput. Chem.*, 2003, **24**, 1624–1636.
43 R. L. Henderson, *Phys. Lett.*, 1974, **49**, 197–198.
44 M. Praprotnik, L. Delle Site and K. Kremer, *Phys. Rev. E*, 2006, **73**, 066701.
45 S. Jain, S. Garde and S. K. Kumar, *Ind. Eng. Chem. Res.*, 2006, **45**, 5614–5618.
46 H. Wang, C. Junghans and K. Kremer, *Eur. Phys. J. E*, 2009, **28**, 221–229.
47 J. D. McCoy and J. G. Curro, *Macromolecules*, 1998, **31**, 9362–9368.
48 B. Hess, S. Leon, N. van der Vegt and K. Kremer, *Soft Matter*, 2006, **2**, 409–414.
49 G. Santangelo, A. Di Matteo, F. Muller-Plathe and G. Milano, *J. Phys. Chem. B*, 2007, **111**, 2765–2773.
50 F. Thorpe, J. Zhou and G. A. Voth, *J. Phys. Chem. B*, 2008, **112**, 13079–13090.
51 X. Chen, P. Carbone, G. Santangelo, A. Di Matteo, G. Milano and F. Muller-Plathe, *Phys. Chem. Chem. Phys.*, 2009, **11**, 1977–88.
52 D. Stueber, T. Y. Yu, B. Hess, K. Kremer and J. Schaefer, submitted to J. Chem Phys., 2009.
53 B. Hess, C. Peter, T. Ozal and N. F. A. van der Vegt, *Macromolecules*, 2008, **41**, 2283–2289.
54 T. Ozal, C. Peter, B. Hess and N. F. A. van der Vegt, *Macromolecules*, 2008, **41**, 5055–5061.
55 B. Hess and N. F. A. Van der Vegt, *Macromolecules*, 2008, **41**, 7281–7283.
56 O. Hahn, D. A. Mooney, F. Muller-Plathe and K. Kremer, *J. Chem. Phys.*, 1999, **111**, 6061–6068.
57 S. Leon, N. F. A. van der Vegt, L. Delle Site and K. Kremer, *Macromolecules*, 2005, **38**, 8078–8092.
58 V. A. Harmandaris and K. Kremer, *Soft Matter*, 2009, DOI: 10.1039/b905361a.
59 V. A. Harmandaris, N. P. Adhikari, N. F. A. van der Vegt, K. Kremer, B. A. Mann, R. Voelkel, H. Weiss and C. C. Liew, *Macromolecules*, 2007, **40**, 7026–7035.
60 D. Fritz, C. R. Herbers, K. Kremer and N. F. A. van der Vegt, *Soft Matter*, 2009, DOI: 10.1039/b911713j.
61 E. Lyman, F. M. Ytreberg and D. M. Zuckerman, *Phys. Rev. Lett.*, 2006, **96**, 028105.
62 D. M. Zuckerman, *Coarse-Graining of Condensed Phase and Biomolecular Systems*, Chapman and Hall/CRC Press, Taylor and Francis Group, 2008.
63 M. Christen and W. F. van Gunsteren, *J. Chem. Phys.*, 2006, **124**, 154106.
64 M. Praprotnik, L. Delle Site and K. Kremer, *J. Chem. Phys.*, 2005, **123**, 224106.

65 M. Praprotnik, L. Delle Site and K. Kremer, *Annu. Rev. Phys. Chem.*, 2008, **59**, 545–571.
66 R. Delgado-Buscalioni, K. Kremer and M. Praprotnik, *J. Chem. Phys.*, 2008, **128**, 114110.
67 H. J. Limbach, A. Arnold, B. A. Mann and C. Holm, *Comput. Phys. Commun.*, 2006, **174**, 704–727.
68 D. Van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *J. Comput. Chem.*, 2005, **26**, 1701–1718.