PCCP

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2025, **27**, 17944

Calculation of consistent neutron-weighted total structure factors from coarse-grained simulation data

Hima Bindu Kolli,*[a] Guadalupe Jiménez-Serratos,[b] James Doutch,[a] Tristan G. A. Youngs[a] and Thomas F. Headen*[a]

The combined use of neutron scattering experiments with molecular simulation is increasingly being used to study multi-scale structures in molecular biology and soft matter physics. Small-angle neutron scattering (SANS) can provide experimental data at the length scale from 1 to 100's of nm, an order of magnitude larger than the typical atomistic simulations. In this context, coarse-grained (CG) simulation can be used to reduce computational costs, explore system polydispersity, and overcome slow dynamics. The mathematical expression to calculate SANS curves from molecular models is well defined for atomistic systems, but further approximations are needed to analyse CG models, where the atomistic resolution is lost. Here, we present the MuSSIC tool, which is a code to compute the neutron-weighted total structure factor, $F_{CG}(Q)$, directly from CG simulation trajectories, based on the methodology proposed by Soper and Edler [*Biochim. Biophys. Acta*, 2017, **6**, 1861]. We validate the approximations by comparing the results against the atomistic pseudo-CG data to decouple force-field effects. We demonstrate the scientific usefulness and understanding provided by the code, by comparison of CG simulations to the experimental scattering data for archetypal soft matter systems, SDS and CTAB solutions. We were able to use the marked differences with experimental SANS data to give a detailed understanding of the appropriateness of the CG simulation methodologies used for predicting structure. This forms a first step towards new approaches in SANS data analysis, particularly in allowing refinement of models against one or more experimental data sources.

Received 11th April 2025,
Accepted 1st August 2025

DOI: 10.1039/d5cp01390a

rsc.li/pccp

1 Introduction

Understanding the structural properties of materials across different length scales remains extremely challenging in spite of decades of research in material and chemical science.[1–4] Small angle neutron scattering (SANS) and neutron total scattering techniques are powerful tools for the characterization of structure in complex, disordered and multi-scale materials, with applications ranging from atomic resolution analysis of individual atoms in liquid systems, through to meso- and macro-scale analysis of complex biological structures, membranes, and assemblies. The use of molecular simulations, as a complementary technique, is of growing importance for these methods.

It is well known that one cannot uniquely assign a 3-dimensional atomistic structure to a particular scattering pattern (in all but the most trivial cases), however, total and small-angle neutron scattering intensity can be calculated from known atomic positions over the course of a simulation trajectory. In the most limited interpretation, this provides a strong structural benchmark that the derived force-field parameters and simulation methodologies do reflect experimental reality. In the cases where there is good agreement between simulation and scattering, the resulting simulation ensemble can be interpreted as a reasonable structure representation of the system in question. While there may be several possibilities of atomic arrangements which give a close match to the same neutron scattering data, the additional constraints on known chemical physics in molecular simulations (density, known bonded molecular structure, force-fields that have reasonable agreement with thermodynamic properties) help reduce the possibilities, and constrain the search for a reasonable structure. Using this approach, the combined use of neutron scattering and molecular simulations has allowed the study of disordered systems like molecular liquids, novel solvents, confined fluids, surfactants, biomolecules and polymers in unprecedented detail.[5–11] For example, methods like empirical potential structure refinement (EPSR), as implemented in the

[a] ISIS Neutron and Muon Source, Rutherford Appleton Laboratory, Harwell Campus, Oxon, OX11 0QX, UK. E-mail: hima-bindu.kolli@stfc.ac.uk, tom.headen@stfc.ac.uk

[b] STFC Hartree Centre, Scitech Daresbury, Warrington WA4 4AD, UK

EPSR code[12–14] and now Dissolve,[15] have been developed to drive the simulation towards matching experimentally measured structure, by comparing the simulations with the experimental scattering data (usually several isotopically distinct datasets), then by means of the feedback of an additional potential based on the difference, the simulation is driven towards matching all available scattering data. The EPSR method has been widely used for interpreting the atomistic structure of liquids and glasses of a wide variety of systems including simple molecular liquids,[16,17] ionic liquids,[18] deep eutectic solvents,[19] multicomponent glasses,[20] biomolecules in solution,[21] fluids in confinement[22] and, at the largest length scale, small micellar systems[23] and biomolecular aggregates,[24] by refining against X-ray and/or neutron total scattering data.

Modern neutron instrumentation, such as the Near and InterMediate Range Order Diffractometer (NIMROD) at the ISIS facility,[25] have the ability to study systems across length scales ranging from the interatomic (<1 Å) through to the mesoscopic (>300 Å) simultaneously. There is therefore a motivation to increase the length scale of the EPSR technique to match the capability of NIMROD to measure atomistic and mesoscale correlations simultaneously. Furthermore, atomistic molecular simulation is increasingly being used to interpret SAS data. Modern SANS instruments, for example SANS2d[26] at the ISIS second target station, offer a wide Q-range and low background, giving experimental data which is more conducive to analysis by simulation based methods.[27]

Computational approaches have been particularly successful for structural studies of large biomolecules, such as proteins in dilute solution, where tools such as SASSIE,[28] have established simulation as a viable analysis method for the lay user, who is not an expert in molecular dynamics. This is a modular framework which contains various elements and operations required to undertake atomistic simulations on biomolecules, including structure building, a number of simulation and minimization modes, and calculation of small-angle scattering curves. This approach has a successful track record in the analysis and interpretation of scattering data for biomolecular systems, for example understanding the conformational space of different components in large complexes such as antibodies, and suggesting plausible conformations for highly flexible biological assemblies.[29–31] However there is an increasing demand to use these tools for understanding more generic soft matter systems, often at high concentration, and where the solvent plays an important role and cannot be ignored in the scattering calculation – therefore system size becomes even more critical. This is particularly the case for concentrated solutions, or in cases where the effect of the hydrating water cannot be ignored where fully atomistic molecular dynamics simulations of soft matter systems and large bio-molecular systems are required. For example, investigating the structural and dynamical features of polymer melts at length scales covering both intermolecular range and local short-range is challenging due to the different relaxation times involved.[32]

Coarse-grained (CG) simulation provides a means for simulating the assembly and interactions of such macromolecular complexes at a reduced level of representation, thereby allowing both longer timescale and larger-sized simulations. In CG simulation, a small number of atoms (typically 3 non-H atoms) are grouped together to form a single CG bead. However, there are certain limitations on increasing the level of coarse-graining due to unphysical bond crossing caused by the softening of the coarse potentials, inaccurate dynamics at interfaces and non-transferability of force field (FF) parameters.[33–35] By obtaining the neutron scattering data directly from CG simulations and comparing it to the experiments, one can perform pair structural analysis and also check the accuracy of the CG potential. But computing neutron scattering from CG simulations is not straightforward as there is a loss in atomic resolution with the definition of the CG beads.

Soper and Edler have developed a preliminary coarse-grained version of empirical potential structure refinement, CG-EPSR, that is potentially applicable to a variety of mesoscale and nanoscale structures.[36] The method closely follows EPSR and involves deriving an empirical interaction CG potential from the scattering data and is applied on reverse aqueous micelle of sodium-dioctyl sulfosuccinate (AOT) and iso-octane, with average CG bead containing ∼200 atoms and radius ∼0.9 nm which is considerably larger than typical CG bead sizes. In this work, the total structure factor $F_{\mathrm{CG}}(Q)$ is computed for the CG simulations by combining the neutron scattering due to the atoms within the single CG bead and scattering due to the bead pairs. The method has not been tested on typical CG bead sizes (3 or 4 carbon atoms per bead) and the intra-molecular bead scattering coming due to the beads that are connected in the molecule is not estimated.

The aim of the current paper is to generalise and verify the neutron scattering calculation method for CG simulations proposed by Soper and Edler in ref. 36 so that it can be used to calculate neutron scattering from any typical CG molecular simulation. The method has been modified to consider much smaller and typical CG bead sizes which are used in standard CG simulations like dissipative particle dynamics or simulations based on MARTINI force fields (≤ 6 carbon atoms per bead) and are bonded. This is done by considering intra- and inter-molecular bead scattering separately, as shown in Section 2.2, allowing efficient calculation of scattering from different isotopically labelled samples.

The key questions we aim to answer with this study are: (i) How well does a CG calculation of scattering work for "typical" CG simulation bead sizes? (ii) What is the effect of bead size and form factor on the calculation accuracy? And (iii) how widely applicable is the calculation method to different soft matter systems? As a final objective, we demonstrate the utility of the code by using it to compare large CG simulations of micellar systems against experimental data. The first version of the 'MuSSIC: Multiscale Simulation Scattering Intensity Calculator'[37] software has been made available on GitHub (**https://github.com/disorderedmaterials/MuSSIC**). All the example files and test systems are included in the user guide and documentation which is provided along with the code.

The paper is organised as follows: first, we include an introduction of the relevant scattering theory and the equations used for the calculation of scattering from CG simulation. This is followed by outlining the simulation and verification methodology used and the results of those validation tests. Finally, we include a demonstration of the utility of the MuSSIC code by calculating the scattering from two coarse-grained surfactant solution simulations (CTAB and SDS) and compare the results to experimental SANS data.

2 Scattering calculation: relevant theory

2.1 Scattering calculation for atomistic simulations

In a neutron scattering experiment, the measured differential scattering cross section (DCS) is obtained after suitable corrections for the beam attenuation, multiple scattering, and inelasticity effects. The interference scattering cross section, or total structure factor, $F(Q)$, is related to the measured DCS *via*

$$\mathrm{DCS} = F(Q) + \sum_{i=1}^{N} c_i b_i^2 \tag{1}$$

where c_i and b_i are the concentration and scattering length of atom type i respectively and the sum is over N number of atom types in the system. It is this interference scattering that contains all the structural information from the experiment, and it is common practice in neutron total scattering experiments to remove the self-scattering background (second term in eqn (1)), to leave just the interference function. Here we follow that practice using the standard formalism as outlined by Keen,[38] noting that this is different to standard SANS formalisms, where the self-scattering is not subtracted and often dealt with through fitting of a constant background in data analysis. We note here that for clarity and to allow negative values of $F(Q)$ to be plotted on a logarithmic scale, we frequently plot $F(Q)$ + 1 as the y-axis to allow for better comparison at higher Q where the scattered signal oscillates around zero.

The total structure factor for a system containing N distinct atom types can be written as the weighted sum of all possible partial pair structure factors $S_{ij}(Q)$.

$$F(Q) = \sum_{i=1, j\geq i}^{N} (2 - \delta_{ij}) c_i c_j b_i b_j \left[S_{ij}(Q) - 1\right] \tag{2}$$

$S_{ij}(Q)$ is the partial structure factor obtained from the spatial correlation between the atom types i and j. The Faber and Ziman definition of partial structure factor $S_{ij}(Q)$ is given as

$$S_{ij}(Q) - 1 = 4\pi\rho \int_0^\infty r^2 \left[g_{ij}(r) - 1\right] \frac{\sin Qr}{Qr} \mathrm{d}r \tag{3}$$

where $g_{ij}(r)$ is the partial radial distribution function of atom types i and j. The product of the compositions c_i and c_j and the scattering lengths b_i and b_j makes the weights matrix element corresponding to partial structure factor $S_{ij}(Q)$ of atom types i and j.

The weights matrix W is calculated for all possible pairs of atom types as given below

$$W_{ij} = (2 - \delta_{ij}) c_i c_j b_i b_j. \tag{4}$$

Here, δ_{ij} is the Dirac delta function which follows

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise.} \end{cases}$$

In atomistic simulations, the partial pair radial distribution functions $g_{ij}(r)$ can be obtained from the atomic positions over the course of many iterations/time steps to obtain an ensemble average. A Fourier transform of $g_{ij}(r)$ yields $S_{ij}(Q)$ and the total structure factor $F(Q)$ can be obtained from the sum of neutron-weighted $S_{ij}(Q)$ as given in eqn (2). Thus an atomistic structure from the simulation can be experimentally validated by comparing $F(Q)$ from simulations with the same obtained from experiments.

For multicomponent systems, a single total structure factor obtained in experiments contains all the partials that one would like to extract individually. The isotopic substitution technique helps to identify the contributions from different partial structure factors (usually hydrogen is replaced by deuterium). Different ratios of isotope give different isotopologues (*i.e.* chemically identical samples where the only change is substituting one or more isotope for another).

The isotopic substitution can be implemented for the calculation of $F(Q)$ from simulations by simply changing the neutron scattering length b_i of the atom type i on the r.h.s of eqn (2) with the scattering length of its isotope. For systems with exchangeable hydrogens (*e.g.* water), the r.h.s of eqn (2) becomes the product of concentration and scattering length of the isotope and its isotope ratio λ. The effective scattering length of hydrogen when substituted for deuterium becomes

$$b_{\mathrm{H}}^{\mathrm{eff}} = \lambda b_{\mathrm{D}} + (1 - \lambda) b_{\mathrm{H}} \tag{5}$$

where b_{H} is the bound coherent scattering length of hydrogen (−3.74 fm), b_{D} is the scattering length of deuterium (6.671 fm), and λ is the fraction of hydrogen that is replaced with deuterium.

Care needs to be taken when calculating the weights matrix (Eq. 4) to understand if the isotopically labelled hydrogens are chemically exchangeable (*e.g.* –OH or –NH) or not (*e.g.* –CH). For example in water a 1 : 1 mix of H_2O and D_2O results in a statistical mixture of H and D across all molecules, however in a 1 : 1 mixture of benzene and d_6-benzene, there is no exchange, and so molecules are either fully deuterated or fully hydrogenated. Therefore, the scattering weight coefficient in eqn (2) is different in each case. To account for this difference, the total $F(Q)$ can be separated into two parts. The structure factor is computed for atoms within the same molecule $F^{\mathrm{intra}}(Q)$ and a separate structure factor is computed for unbound atoms $F^{\mathrm{inter}}(Q)$. Full details of this calculation, with examples, are given in the SI Section S1.

2.2 Scattering calculation for coarse-grained simulations

For a coarse-grained simulation, we swap several atoms (usually ≈ 3 non-H atoms) and as a consequence we lose the local positional information of these atoms. We therefore expect to lose fidelity of the calculated scattering at higher Q as the cost for a substantial increase in both simulation and scattering calculation simplicity and therefore speed. To achieve this we can no longer think in terms of scattering from atoms of a certain scattering length, but must instead consider scattering from a bead with a certain size and distribution of scattering length density. We define the scattering length density distribution $n_s(r)$ of CG bead type 's' from the distribution of atoms within the bead as shown in Fig. 1.

$$n_s(r) = \sum_i \left[N_i^{(s)} b_i\right] \rho(r_s) \tag{6}$$

where N_i^s is the number of atoms of atom type i within bead type s, b_i is the scattering length of atom type i and $\rho(r_s)$ is the density distribution of the atoms for bead type 's'.

By analogy with the atomistic case (eqn (1) and (2)) the total differential scattering cross-section for a CG system containing M distinct CG bead types can be written as[36]

$$\mathrm{DCS} = \frac{1}{N_\mathrm{b}} \sum_s^M c_s \sum_i N_i^{(s)} b_i^2 + F_\mathrm{CG}(Q) \tag{7}$$

The first term indicates the single-atom scattering and the total structure factor for coarse grain systems, $F_\mathrm{CG}(Q)$, is the sum of single-bead scattering $F_\mathrm{CG}^\text{single-bead}(Q)$ and cross-bead scattering $F_\mathrm{CG}^\text{cross-bead}(Q)$. N_b is the average number of atoms per bead.

$$F_\mathrm{CG}(Q) = F_\mathrm{CG}^\text{single-bead}(Q) + F_\mathrm{CG}^\text{cross-bead}(Q). \tag{8}$$

$F_\mathrm{CG}^\text{single-bead}(Q)$ is the scattering from all possible pairs of atom types within the single bead s, obtained by multiplying the total effective scattering length with form factor $f_s(Q)$, which represents the Q dependent scattering from the variation of scattering length density within a bead.

$$\begin{aligned} F_\mathrm{CG}^\text{single-bead}(Q) &= \frac{1}{N_\mathrm{b}} \sum_s^M c_s \left[\sum_i N_i^{(s)}\left(N_i^{(s)} - 1\right) b_i^2 \right. \\ &\quad \left. + 2\sum_i N_i^{(s)} b_i N_j^{(s)} b_j \right] f_s^2(Q). \end{aligned} \tag{9}$$

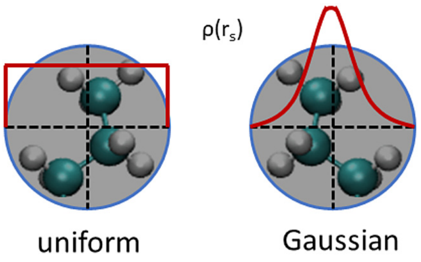


Fig. 1 Schematic showing the uniform and Gaussian density distributions used for the definition of form factor $f_s(Q)$ for bead type 's'.

which can also be written as

$$\begin{aligned} F_\mathrm{CG}^\text{single-bead}(Q) &= \frac{1}{N_\mathrm{b}} \sum_s^M c_s \left[\sum_i \sum_{j \geq i} (2 - \delta_{ij}) N_i^{(s)} b_i N_j^{(s)} b_j \right. \\ &\quad \left. - \sum_i N_i^{(s)} b_i^2 \right] f_s^2(Q). \end{aligned} \tag{10}$$

The extra terms in $F_\mathrm{CG}^\text{single-bead}(Q)$, with respect to the atomistic case, take care of the fact that each bead may contain more than one atom of any given type. Since the atoms in each bead are assumed to be distributed statistically through the bead according to the same $\rho(r_s)$ for all atom types their relative positions within the bead are uncorrelated. The density distribution and the form factor used for the current study are discussed in Section S3 of the SI.

The atoms within the bead are treated similarly as in intramolecular scattering for atomistic simulations. Therefore for non-exchangeable hydrogens, and given an isotope ratio λ, the above equation changes to

$$\begin{aligned} &F_\mathrm{CG}^\text{single-bead}(Q) \\ &= \frac{1}{N_\mathrm{b}} \sum_s^M c_s \left\{ (1 - \lambda). \right. \\ &\quad \times \left[\sum_i \sum_{j \geq i} (2 - \delta_{ij}) N_i^{(s)} b_{\mathrm{nat},i} N_j^{(s)} b_{\mathrm{nat},j} - \sum_i N_i^{(s)} b_{\mathrm{nat},i}{}^2 \right] \\ &\quad \left. + \lambda \left[\sum_i \sum_{j \geq i} (2 - \delta_{ij}) N_i^{(s)} b_{\mathrm{iso},i} N_j^{(s)} b_{\mathrm{iso},j} - \sum_i N_i^{(s)} b_{\mathrm{iso},i}{}^2 \right] \right\} f_s^2(Q) \end{aligned} \tag{11}$$

where $b_{\mathrm{nat},i}$ is the scattering length of atom type i and $b_{\mathrm{iso},i}$ is the scattering length of the isotope of atom type i.

$F_\mathrm{CG}^\text{cross-bead}(Q)$ is the scattering from all possible pairs of bead types, which is obtained by analogy to the atomistic case (see eqn (2)), where this time the "bead scattering length" is obtained by multiplying the sum of the scattering lengths of atoms in the bead s with its form factor $f_s(Q)$

$$\begin{aligned} F_\mathrm{CG}^\text{cross-bead}(Q) &= \frac{1}{N_\mathrm{b}} \sum_s^M \sum_{t \geq s}^M (2 - \delta_{st}) c_s c_t \left[\sum N_i^{(s)} b_i\right] f_s(Q) \\ &\quad \times \left[\sum N_i^{(t)} b_i\right] f_t(Q)[H_{st}(Q) - 1] \end{aligned} \tag{12}$$

Here $H_{st}(Q)$ is the partial structure factor obtained from the spatial correlation between the bead types s and t and is obtained following the Faber and Ziman definition shown in eqn (3). In similarity to the scattering for atomistic models, the total structure factor is divided into $F_\mathrm{CG}^\text{intra}(Q)$ and $F_\mathrm{CG}^\text{inter}(Q)$:

$$F_\mathrm{CG}^\text{cross-bead}(Q) = F_\mathrm{CG}^\text{intra}(Q) + F_\mathrm{CG}^\text{inter}(Q) \tag{13}$$

Intra and inter-molecular scattering curves are obtained following the same equation as in the scattering from atomistic simulations. The only difference is in the computation of weights matrices for partial structure factors $H_{st}(Q)$ which is outlined in detail in Section S2 of the SI.

We also need to consider how to distribute the scattering length density over a bead, as a function of its radius. We describe this in detail in SI Section S3. In summary we consider both a uniform and Gaussian distribution (see Fig. 1, which result in different bead form factors $f(Q)$).

Finally, it is important to note that the lower limit of the Q-range, $Q_{\min}$, for obtaining reliable scattering data depends on the simulation box size L. The maximum distance between the beads in the $g(r)$ calculation is limited to half of the box due to the periodic boundary conditions. It is therefore reasonable to set $Q_{\min}$ as equivalent to the maximum correlation in r-space, *i.e.* $Q_{\min} = 2\pi/(L/2)$.

3 Verification and simulation methodology

3.1 Verification of CG scattering calculation method against an atomistic benchmark

In order to test the accuracy of eqn (7) as a method for calculation of neutron scattering from CG models, *atomistic* molecular dynamics simulations are first performed on selected systems. The total neutron scattering $F(Q)$ is obtained following the equations given in Section 2.1 and is taken as benchmark data. A module in the MuSSIC code has been used to do this and the calculation has been validated by comparing it with the same obtained from the *Dissolve* software package.[15] Pseudo-CG trajectories are generated by replacing a group of atoms with a bead located at the geometric center of the group of atoms. In Fig. 2, we show an example of the conversion.

The MuSSIC code is used to compute $F_{\text{CG}}(Q)$ from the pseudo-CG trajectory following the equations shown in Section 2.2. $F_{\text{CG}}(Q)$ is then compared against the atomistic benchmark data $F(Q)$ to test the accuracy of the calculation. Fig. 3 shows the workflow followed for the validation tests. This comparison test is repeated for different pseudo-CG trajectories generated using different CG mapping models, but not by performing a CG simulation. This allows us to compare $F_{\text{CG}}(Q)$ with $F(Q)$ for the same underlying structure, just with different levels of spatial resolution. The accuracy of the method is tested by comparing the $F_{\text{CG}}(Q)$ with reference data $F(Q)$ from the atomistic trajectory. We are therefore able to remove any difference that occurs between atomistic and CG representations due to differences in the simulation force-fields, affording a direct comparison and assessment of the quality of the CG calculation method against and atomistic one. Validation tests have been performed following the workflow shown in Fig. 3 on a polyamide-66 melt and a solution of C_{10} TAB surfactant in water for different CG mapping models and the results are discussed in Section 4.1. Three different pseudo-CG mapping schemes were used for each system as shown in Fig. 4 and 5. Details for the setup of the atomistic simulation are given in Section S4 of the SI. Fig. 6a shows the atomistic representation of a configuration of polyamide-66 alongside the same with ∼4 carbon atoms per bead (Fig. 6b) representation.

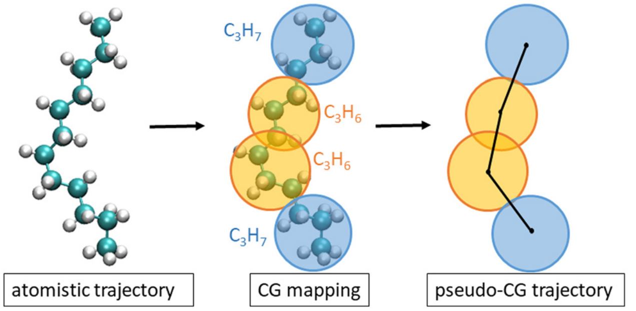


Fig. 2 Atomistic trajectories are converted to pseudo-CG trajectories by replacing a group of atoms with a bead according to a CG mapping model. Different colors indicate different atom or bead types in each representation.

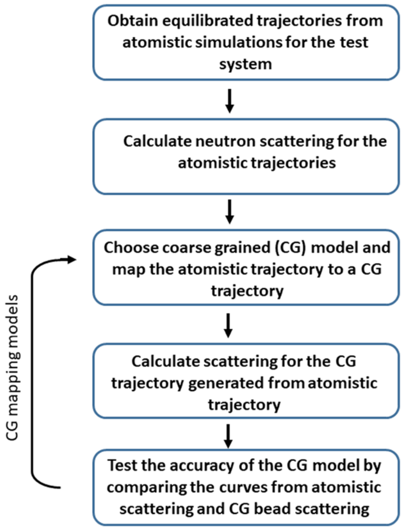
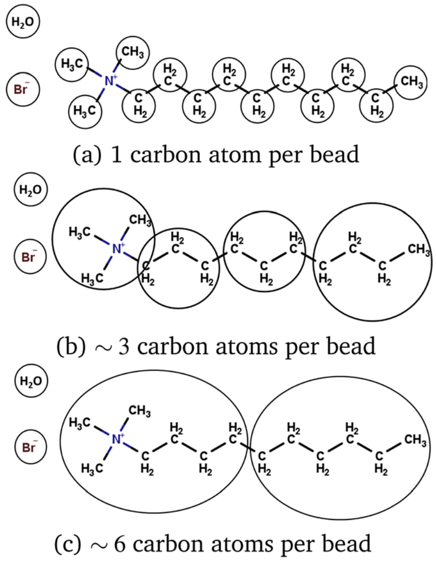

Obtain equilibrated trajectories from atomistic simulations for the test system
Calculate neutron scattering for the atomistic trajectories
Choose coarse grained (CG) model and map the atomistic trajectory to a CG trajectory
Calculate scattering for the CG trajectory generated from atomistic trajectory
Test the accuracy of the CG model by comparing the curves from atomistic scattering and CG bead scattering
CG mapping models

Fig. 3 Verification method followed for testing the efficacy of the CG scattering calculation.

3.2 Overview of coarse-grained simulation methodology

Having validated the code using atomistic simulation, we then aim to use the calculation method to compare genuine CG simulations to experimental SANS data. We do this using two different micellar systems each with a different CG simulation methodology. Firstly we simulate SDS micelles in water using hybrid particle-field molecular dynamics simulations (hPF-MD).[39,40] These simulations were performed using the OCCAM software.[41–44] In this hybrid approach, the intramolecular interactions are treated by a standard molecular Hamiltonian, and the intermolecular forces are described by density fields. Electrostatics are implemented as an additional external field obtained by a modified particle–mesh Ewald procedure.[45] The OCCAM software has been developed based on hPF-MD method and its performance has been studied thoroughly and documented in ref. 46. Full details of the simulation methodology and setup are provided in the SI Sections S5.1 and S5.2 and the overview of the CG beads used in the model

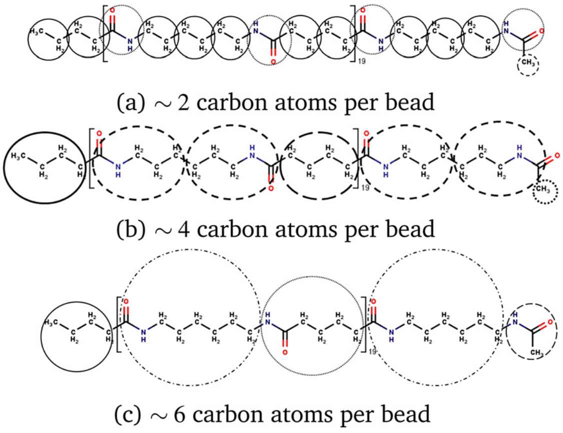


(a) 1 carbon atom per bead

(b) ∼3 carbon atoms per bead

(c) ∼6 carbon atoms per bead

Fig. 4 Mapping between atomistic and CG models of C_{10}TAB. The circles denote which atoms are grouped into CG beads. Figure shows (a) a 14-bead surfactant with ∼1 carbon atoms per bead model, (b) a 4-bead surfactant with ∼3 carbon atoms per bead model and (c) a 2-bead surfactant with ∼6 carbon atoms per bead model with an average bead radius of ∼0.7, ∼1.5, and ∼2.5 Å, respectively.

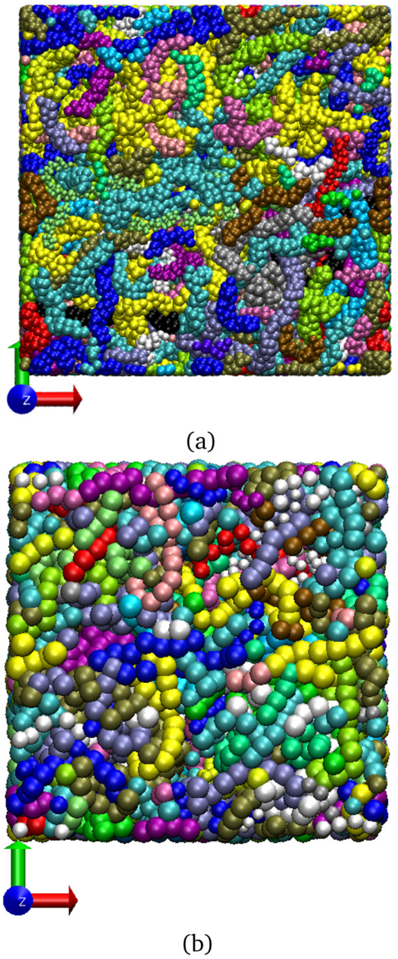


(a) ∼2 carbon atoms per bead

(b) ∼4 carbon atoms per bead

(c) ∼6 carbon atoms per bead

Fig. 5 Mapping between atomistic and CG models of polyamide-66. The circles denote which atoms are joined into CG beads. Figure shows (a) a chain with 141 beads using ∼2 carbon atoms per bead model, (b) a chain with 100 beads using ∼4 carbon atoms per bead model and (c) a chain with 41 beads using and ∼6 carbon atoms per bead model with an average bead radius of ∼1.3, ∼2.5, and ∼4.5 Å, respectively.



(a)

(b)

Fig. 6 Snapshots showing the (a) atomistic and (b) ∼4 carbon atoms per bead representation of polyamide-66 system.

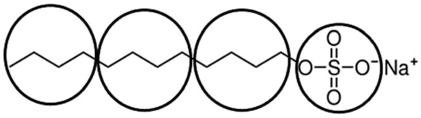


Fig. 7 Structure of SDS and CG mapping model used in hPF-MD simulations.

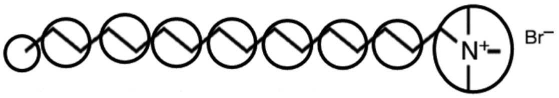


Fig. 8 CG representation of C_{16}TAB surfactant for DPD simulations.

are shown in Fig. 7, noting that water is treated as having 4 molecules per bead, and sodium cations in a bead with 4 water molecules. We further note that the fact that each water bead contains multiple molecules has important implications for the ability of the model to correctly predict local water structure $g(r)$, as is discussed in detail in Section 4.2.

Secondly dissipative particle dynamics (DPD) simulations[47–49] were performed using LAMMPS v18Mar2018[50] to study the formation of C_{16}TAB micelles in water. The coarse-grained model is based on the CTAC model presented in ref. 51, where the surfactant is a flexible chain of 9 beads representing $C_{16}TA^+$ as $[(CH_3)_3N^+CH_2][CH_2CH_2]_7[CH_3]$, and the counter ion Br^- is contained in a $[Br^-(H_2O)_2]$ bead as shown in the Fig. 8. For the water model, two molecules are implicit per DPD segment, $(H_2O)_2$. The non-bonded interactions are modelled *via* soft-repulsive potentials with parameters taken from ref. 51–53. Full details of the simulation methodology and setup are provided in the SI Sections S5.3 and S5.4.

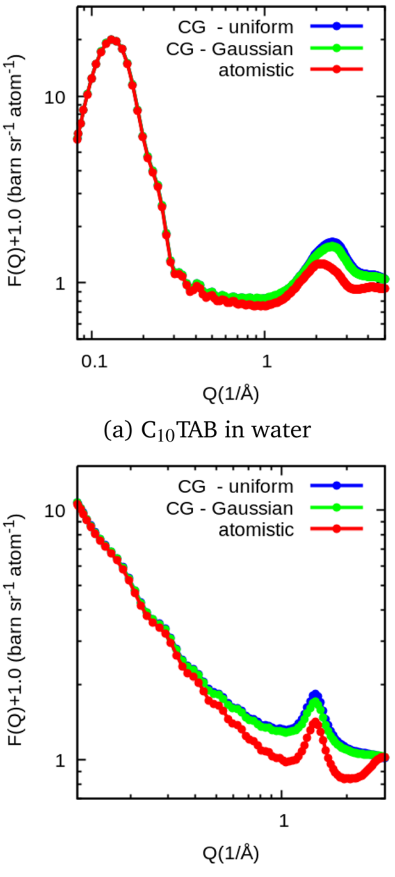
4 Results and discussion

In the first part of this section we demonstrate the accuracy of the MuSSIC neutron scattering calculation method as described in Section 2.2. Validation tests on three systems are performed by following the verification method given in Section 3.1 and compared to the scattering obtained from the reference atomistic simulations (details given in Section S4 of SI). In the second part we demonstrate the utility of the method by calculating the scattering directly from two different CG simulations of soft matter systems.

4.1 Validation of the method

4.1.1 Validation test 1: scattering length density distribution and form factor. Gaussian and uniform scattering length density (SLD) distributions are used to obtain the form factor for the CG beads defined by the mapping shown in Fig. 4 and 5. Fig. 9 shows $F_{CG}(Q)$ compared against atomistic $F(Q)$ for polyamide-66 and C_{10}TAB surfactant in water. The CG mapping in both cases is achieved by replacing roughly 2 or 3 carbon atoms with a CG bead.

Fig. 9 shows that the difference between the Gaussian and uniform scattering length density distributions is very small, and negligible at low Q values. However, the Gaussian distribution shows slightly better agreement for both polyamide-66 and C_{10}TAB surfactant in water cases shown in Fig. 9a and b. Therefore, the Gaussian distribution has been used throughout the rest of the study with the width of the Gaussian, $\gamma_s = 0.51R_s$ (where R_s is the uniform sphere radius) for all the test cases as explained in SI Section S3.



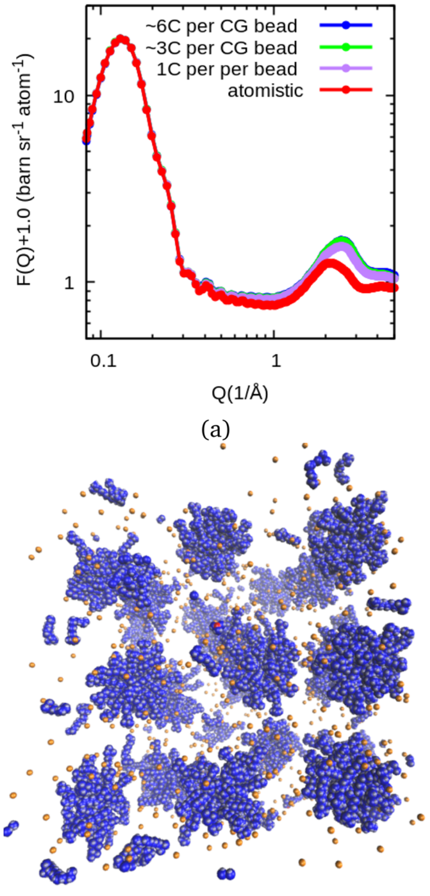
(a) C_{10}TAB in water

(b) polyamide-66

Fig. 9 $F_{CG}(Q)$ computed using Gaussian and uniform distribution of scattering length density and compared against $F(Q)$ calculated from the atomistic trajectories of (a) C_{10}TAB surfactant in water and (b) polyamide-66.

The radius of the CG bead (R_s), for the bead type s is a variable in the calculation which affects the form factor of the bead type f_s and thus affects the total scattering. For the test cases shown in Fig. 9, the radius of the CG bead is estimated from the distribution of atoms defining the bead by using the root mean square deviation (rmsd) of the distances of individual atoms (composing the bead) from the geometric center of the bead. In addition, we also estimated the radius of the beads from bead connectivity (beads that are bonded within a molecule) in the pseudo-CG trajectory. We found that the radius calculated from an rmsd calculation showed a clearer intermolecular structure peak closer to the atomistic calculation, although with a higher background, making the overall residual greater. Given this, we choose to use the RMSD calculation as the standard method for calculating the bead size as it is more universal and has the benefit of simplicity of implementation. Further details of these tests, allowing with further tests of the affect of bead size on the calculation are given in the SI Section S6.

4.1.2 Validation test 2: C_{10}TAB surfactant in water. Fig. 10a shows the total neutron scattering $F_{CG}(Q)$ and $F(Q)$ computed for C_{10}TAB surfactant in D_2O. Comparison plots against $F(Q)$ from the reference atomistic simulations are made for three different CG trajectories which are generated using CG mapping schemes shown in Fig. 4. Fig. 10b shows the snapshot of the stable micelles of C_{10}TAB surfactant (water has been removed for visual clarity). The absolute and relative differences between $F_{CG}(Q)$ and $F(Q)$ as a function of Q are plotted in Fig. S12 and S15 in Section S7 of the SI, along with a further discussion of the differences. Examining the calculated scattering we observe that the first (low Q) peak in the Fig. 10a is due to the structure factor of the micelles, *i.e.* the interactions of the micelles formed by C_{10}TAB surfactants in water (inter-micelle structure factor) whereas the second higher Q peak ($Q \approx$ 2–3 Å^{-1}) originates from the local order in bulk water.[17] There is a good match of $F_{CG}(Q)$ to the scattering from an atomistic simulation $F(Q)$ at low Q, with a definite mismatch at higher Q as expected due to the loss of atomistic resolution in the CG representation. However it is perhaps interesting to observe that this higher Q peak is observed at all in the CG-representation, albeit shifted to a higher Q. This shift is likely due to the loss of local structural information when coarse-gaining the system *i.e.* moving from a water molecule defined by 3 points (with roughly equal scattering length in the case of D_2O) to a spherically symmetric bead centred on the geometric centre of the molecule. Fig. S10 in the SI shows that this shift is dependent on the bead radius used, *i.e.* how spread out the scattering length density is. If the radius is doubled from the radius obtained from the rmsd the peak positions match well in



(a)

(b)

Fig. 10 (a) CG scattering $F_{CG}(Q)$ calculated by overlaying CG beads on to 1, ∼3, and ∼6 carbon atoms and compared against scattering from atomistic simulations of C_{10}TAB surfactants in water. (b) Snapshot showing stable spherical micelles of C_{10}TAB surfactants in water at the end of 90 ns long simulation. colour code: $C_{10}TA^+$ – blue; Br^- – orange and water has been omitted for visual clarity.

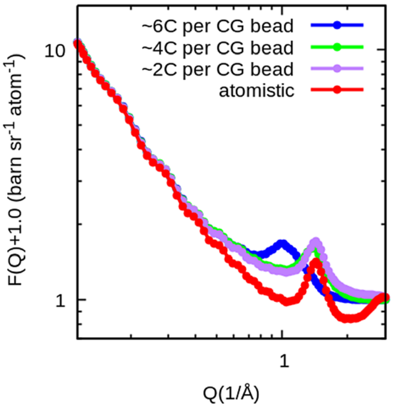


Fig. 11 $F_{CG}(Q)$ computed for polyamide-66 and compared with $F(Q)$ using the CG mappings in Fig. 5 (a) ∼2 carbon atoms per bead (b) ∼4 carbon atoms per bead and (c) ∼6 carbon atoms per bead.

$F(Q)$ but at the expense of broadening the peak and giving a reduced match to atomistic scattering in the region of 1 $Å^{-1}$. Finally we note that Fig. 10 shows that there is very little difference in $F_{CG}(Q)$ due to the different CG bead mapping (see Fig. 4) used for surfactant. This is likely due to the same water bead size has been used for all systems. Changing the size of the water CG representation was not possible due to the complexity of appropriately mapping more than one molecule onto a CG bead.

4.1.3 Validation test 3: polyamide-66. This system is a dense polymer melt composed of 192 chains each having 765 atoms. $F(Q)$ is calculated from the atomistic trajectory considering 50% deuterated chains using the MuSSIC code. Pseudo-CG trajectories are generated using the mappings presented in Fig. 5. Note that the hydrogens on the polyamide-66 chain are not exchangeable, meaning that chains are either fully hydrogenated or deuterated. Following the workflow given in Fig. 3, $F_{CG}(Q)$ has been computed for CG trajectories and compared against the atomistic benchmark, $F(Q)$. The results are shown in Fig. 11, with plots of the absolute and relative differences shown in Fig. S13 and S14 in the SI Section S7.

Overall Fig. 11 shows a good agreement in the low Q region for all bead sizes. As expected, with an increase in the bead size, there is an increasing divergence between the atomistic and CG scattering at higher Q, due to the loss of spatial resolution. Of particular note is the abrupt change in the position of the high Q scattering peak as the CG representation moves from 4 to 6 beads, which is explored in more detail in the next section. A more detailed breakdown of scattering calculation into $F^{\text{intra}}_{\text{CG}}(Q)$ and $F^{\text{inter}}_{\text{CG}}(Q)$ is given in Section S2 of the SI showing that the differences with bead size primarily come from the intermolecular scattering in the high Q region.

4.1.4 Discussion on validation tests. Overall the differences between scattering calculated from an atomistic simulation and a CG representation are small at low Q, where most structural information is measured on SANS instruments. Relative differences for 4C beads are less than 4% for CTAB and less than 13% for PA66 for Q = 0.5 $Å^{-1}$, decreasing to 0.3% and 1% respectively at Q = 0.1 $Å^{-1}$. However, it is clear that at high Q the information loss in the atomic positions in bead description has manifested in wider $F_{CG}(Q)$ peaks. As expected for most cases the accuracy of the method increases with decreasing Q, and the difference to atomistic calculation increases with an increase in the CG bead size. In particular, there is a clear deviation in the difference for the step up to the largest bead size of ∼6C for both cases – at high Q in the PA66 case, and at low Q for CTAB case (noting that the relative error at low Q is less due to the high level of scattering as shown in Fig. S15 of SI).

Exploring the ∼6C case of PA66 shown in the Fig. 11 in more detail, we see a shift in the peak at (∼1 $Å^{-1}$) to lower Q value. This change in the peak shape and position is not seen in ∼2C and ∼4C cases, so what is it about moving up to ∼6C that causes this discontinuous change? To understand this further, we plot partial pair correlation functions $g_{xy}(r)$ for the CG beads of the polyamide chain. Here x and y indicate CG bead types. Fig. 12 shows the $g_{xy}(r)$ plotted for C2–C2, C4–C4 and C6–C6

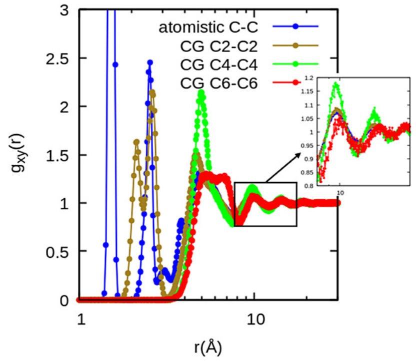


Fig. 12 Partial pair correlation function $g_{xy}(r)$ plotted for the pairs of atom types C–C from atomistic simulations, bead type pairs C2–C2 (∼2 carbon atoms per bead), C4–C4 (∼4 carbon atoms per bead) and C6–C6 (∼6 carbon atoms per bead) from pseudo-CG trajectories. Inlet shows the zoomed picture in the long range.

bead type pairs for ∼2C, ∼4C and ∼6C cases respectively. For comparison, $g_{CC}(r)$ is plotted from atomistic simulations. Obviously, there is a loss of the low r peaks as we move to larger beads, however up to ∼4C beads, the larger r peaks overlap, indicating a faithful representation of the overall structure. However, there is a clear change in the structure on moving to the 6C case with a new peak appearing at ∼7 Å. Furthermore, there is a clear trend of the longer distance peaks shifting to higher r (shown in the inset of the Fig. 12). We can further understand this difference by separating the $g(r)$ into intra and inter-molecular components, as shown in Fig. S16 of the SI. The peak positions remain less affected in inter-molecular correlation shown in Fig. S16b of the SI, while the intra-molecular correlation (Fig. S16a of the SI) shows a clear shift in the peak for ∼6C case. This analysis clearly indicates that the disagreement for the 6C bead case stems from the reduced detail and correspondence of that CG representation with the underlying structure, resulting in a clear difference in the high Q scattering data.

4.2 Use cases: comparing CG simulations to SANS data

From the above, we can now use the calculation method with confidence that discrepancies between an atomistic and CG calculation are small where we are using typical CG simulation bead sizes (3–4 large atoms) and in the typical SANS Q-range ($Q < 0.5$ Å^{-1}). Therefore any significant discrepancy between the experimental and simulated scattering is due to the simulation not correctly representing the experimentally measured structure, rather than an issue with the calculation method itself.

To further understand and exemplify the calculation method, the MuSSIC code has been used on two CG simulations and the outputs compared to experimental neutron scattering (SANS) data, allowing new physical insight into the ability of CG force-fields and methods to predict structure. SANS data were collected on the SANS2d instrument,[26] ISIS Neutron and Muon Source, using source to sample and sample to detector distance of 4 m and circular final (A2) aperture size of 8 mm. The time of flight method was used, utilising wavelengths of 1.75–16.5 Å. Samples were placed in quartz cuvettes in a temperature controlled sample changer at 25 °C. Data were normalised and corrected to 1-D curves using Mantid software.[54,55]

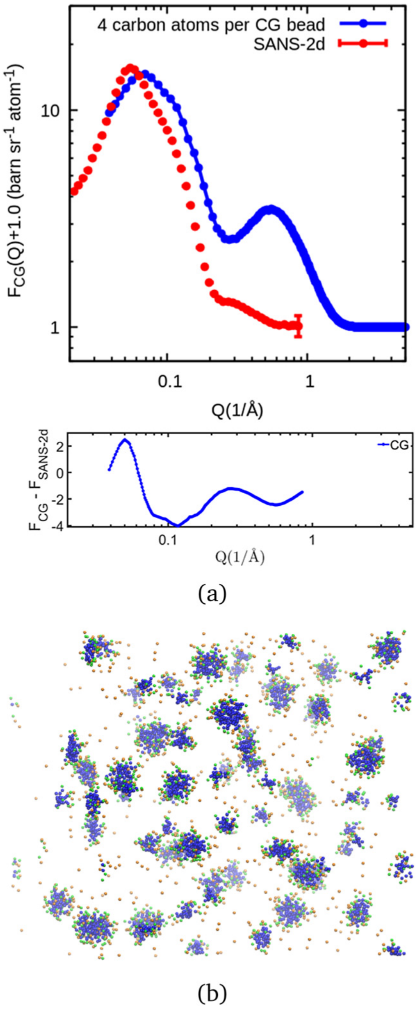


Fig. 13 (a) $F_{CG}(Q)$ calculated for the equilibrated structures of 60 mM of SDS in water and compared to the data from SANS-2D (b) Snapshot taken at the end of 6 μs long hybrid particle-field molecular dynamics simulations. Blue indicates SDS chains and orange indicates Na$^+$ ions in water (water is removed for visual clarity).

4.2.1 SDS surfactant in water. Coarse-grained trajectories of 60 mM SDS surfactants in water are obtained from hybrid particle-field molecular dynamics simulations (hPF-MD) using OCCAM (simulation details are given in Sections S5.1 and S5.2 of the SI). Fig. 13a shows the neutron scattering $F_{CG}(Q)$ obtained for SDS surfactants in D_2O and compared to SANS-2d data (a comparison presentation of the data without addition of one to $F(Q)$ is shown in SI Section S11). Fig. 13b shows the snapshot of spherical micelles of SDS surfactants taken at the end of 6 μs long run. The slope and the peak position of the scattering curve $F_{CG}(Q)$ match reasonably well with the experiments in the Q region (0.03 Å^{-1} < Q < 0.3 Å^{-1}) with larger relative differences seen at higher Q, due to a secondary intermediate Q peak at ≈0.5 Å^{-1} in simulations, the causes of this peak will be discussed in Section 4.2.3.

To further understand the differences between the experiment and CG simulation, we calculated the aggregation properties in this system, such as shape, size, and distribution of micelles, directly from SANS data. This was performed by using model fits in SasView,[56] with the best fit obtained for a spherical model of micelle with radius ∼20 Å and a net charge of ∼14*e*. The Hayter–Penfold Rescaled Mean Spherical Approximation (RMSA) structure factor[57,58] was used to obtain the inter-particle structure factor $S(Q)$. This structure factor was chosen, as it can be used to describe interparticle effects between charged particles. This is a well established method for surfactant systems at the concentrations studied and without counter-ion. A simpler hard sphere interaction would not fit well in these cases. The effective radius of the micelle is calculated by fitting the form factor $P(Q)$ combined with $S(Q)$ as $[P(Q)\cdot S(Q)]$. The average aggregation number $\langle N_{\text{agg}}^{\text{exp}}\rangle$ is obtained from the effective radius of the micelle using (volume$_{\text{micelle}}$/volume$_{\text{surfactant}}$) which is found to be ∼60 surfactants per micelle in SDS case.

The average aggregation number $\langle N_{\text{agg}}^{\text{sim}}\rangle$ was obtained from CG simulations by counting the number of micelles and the molecules in each micelle for each configuration.[59] For this, molecules forming a micelle are identified using a cutoff distance of 8 Å (1.7 times the bead diameter) using the linked list algorithm. The cutoff distance corresponds to the first minimum of the radial distribution function between the head bead of the surfactant and Na^+ ion. The geometric center of the micelle is obtained from the molecules forming it. The radius of the micelle is then obtained from the root mean square deviation of the beads from the center of the micelle.

The average radius of the micelle is found to be ∼23 Å and the average aggregation number $\langle N_{\text{agg}}^{\text{sim}}\rangle$ is found to be ∼28 surfactants per micelle. Though the average radius of the micelle from CG simulations matches closely with the value obtained from SasView analysis of the experimental SANS data, the average aggregation number $\langle N_{\text{agg}}^{\text{sim}}\rangle$ is found to be lower than $\langle N_{\text{agg}}^{\text{exp}}\rangle$. Using the CG simulation we are able to get a much more detailed analysis of the distribution of aggregate sizes for the system. Fig. 14 shows the probability distribution of micelle size $N_{\text{agg}}^{\text{sim}}$ which was obtained after performing 5 μs simulations. The average aggregation number plotted against simulation time is shown in SI Fig. S17, showing that the simulation is run long enough (6 μs) to see a stable value of N_{agg} and no further fusion of micelles happening. We can see a wide distribution of sizes from monomers to micelles of size $N_{\text{agg}}^{\text{sim}}$ = ∼50 surfactants, having a peak around $N_{\text{agg}}^{\text{sim}}$ ∼ 10. The presence of the few bigger micelles in the simulation box make the average aggregation number $N_{\text{agg}}^{\text{sim}}$ less meaningful in such a case. A similar variation in the size distribution of the micelles is seen at low concentrations of SDS (50 mM) even after 5 μs long hPF-MD simulations (reported in Fig. 8 of the ref. 59) confirming sufficient equilibration time in this study. However, hPF-MD simulations are able to show a more stable large number of monodisperse micelles at higher concentrations of SDS.[40,59]

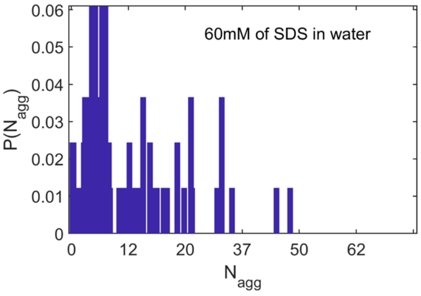


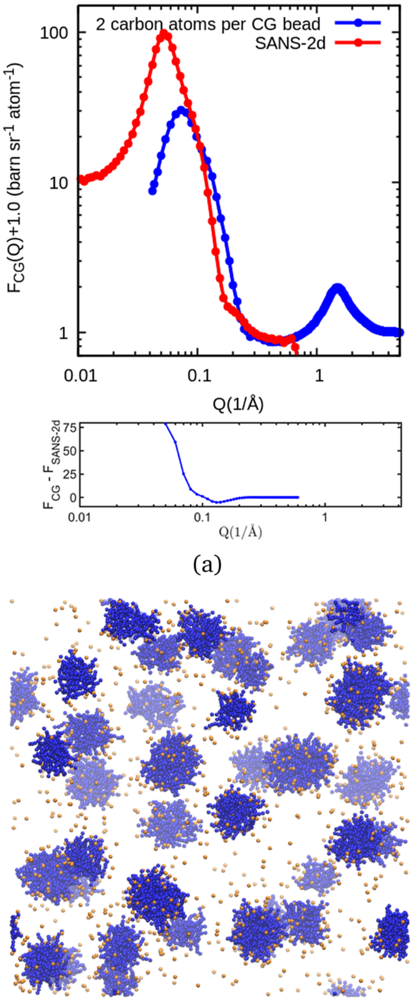
Fig. 14 Probability distribution of the number of SDS micelles *versus* micelle size at 60 mM concentration.

The differences found in the scattering curves between experiments and the simulations are useful in helping our understanding of exactly how the simulation does not match the experiment. In this case, it can be attributed to the fast dynamics in hybrid particle-field potentials resulting in the formation of a few bigger micelles and more smaller aggregates compared to experiments at such low SDS concentrations, resulting in a shift in the micelle structure factor peak to higher *Q*.

4.2.2 C_{16}TAB surfactant in water. Dissipative particle dynamics (DPD) simulations are performed on 100 mM CTAB surfactant system in water for 1.75 μs (simulation details are given in Sections S5.3 and S5.4 of the SI). Fig. 15a shows the calculated SANS data $F_{\text{CG}}(Q)$ obtained for CTAB surfactants in D_2O and compared to SANS-2d data. Fig. 15b shows the final snapshot of spherical micelles of CTAB surfactants at the end of 1.75 μs long DPD simulation. The slope of the curve matches in the region (0.07 Å^{-1} < Q < 0.7 Å^{-1}), however, the low *Q* micelle structure factor peaks is smaller, with the position shifted towards higher *Q* compared with SANS data. The scattering pattern denotes smaller aggregate sizes in the simulations.

A similar SasView analysis of the aggregates has been performed on experimental SANS data as for CTAB, using the Hayter–Penfold Rescaled Mean Spherical Approximation (RMSA) structure factor to obtain the interparticle structure factor. The best fit is obtained for a spherical model of monodisperse micelles with a radius ∼27 Å. The average aggregation number from SANS data $\langle N_{\text{agg}}^{\text{exp}}\rangle$ is found to be ∼135 surfactants per micelle. Simulations give an average radius of the micelle ∼17.8 and an aggregation number of $\langle N_{\text{agg}}^{\text{sim}}\rangle$ ∼42 surfactants per micelle showing the presence of considerably smaller CTAB micelles compared to experiments. The probability distribution of aggregate size $N_{\text{agg}}^{\text{sim}}$, shown in Fig. 16, shows a major proportion of more stable, similar-sized aggregates (∼42 surfactants per micelle) in the simulation box and by calculating the average cluster size as a function of time shown in Fig. S18 in the SI, we found that the equilibrium was reached after 25×10^6 time steps, using a time step of 0.01 in DPD units. We ran 10^7 equilibrium steps more, from which the frames for analysis were taken.

Unlike the SDS case, DPD simulations of CTAB show less polydispersity in the micellar size distribution with no second peak formation. The smaller and shifted peak reflects the smaller aggregates compared to experiments. This has been



(a)

(b)

Fig. 15 (a) $F_{CG}(Q)$ calculated for the equilibrated structures of 100 mM of CTAB in water and compared to the data from SANS-2D instrument (b) Snapshot taken at the end of 1.75 μs long DPD simulations. Blue indicates CTAB chains and orange indicates Br^- ions in water (water is removed for visual clarity).

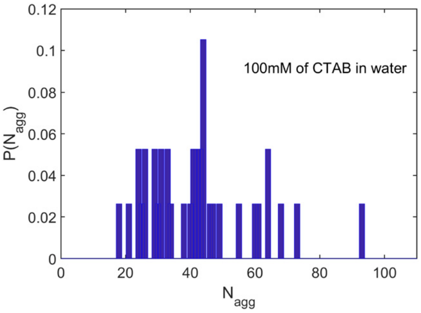


Fig. 16 Probability distribution of number of CTAB micelles *versus* micelle size at 100 mM concentration.

observed in other DPD studies of CTAB micelles.[60] The direct comparison of the simulation to the experimental data provides for a much richer, more detailed analysis than comparing outputs of fitted SAS models, and does not rely on the inherent assumptions in those models (micelle shape, polydispersity models *etc.*).

4.2.3 Discussion on structure prediction from CG simulation. First we consider the differences between the simulation and experiment at higher Q ($Q > 0.3$ Å^{-1}). For DPD simulations of a CTAB solution the peak at 1–2 Å^{-1} is expected due to liquid structural order from D_2O. We note that this is beyond the range of typical SANS data, but it is interesting that the CG representation can still represent this structure with a peak position close to the experimental peak for D_2O at 1.95 Å^{-1} (ref. 17) despite the loss of atomistic level detail in going to a single bead per molecule. In contrast, for the hybrid particle field simulations of SDS there is a peak/shoulder at 0.3–1 Å^{-1} (herein termed as "intermediate Q peak") that is not expected, not in the experimental SANS data and therefore requires further analysis. To understand this further we have calculated the scattering pattern during the micellization process from the start of a simulation where all molecules are randomly placed in the simulation box. The data are plotted at intermediate times while the system is moving towards equilibrium.

Fig. 17a shows the scattering data at 0 ns, 5 ns, 90 ns and at the end of 1.75 μs long DPD simulation of CTAB surfactants in water. A random mix of surfactants in water is used as the initial configuration (0 ns). The low Q peak emerges as micelles start growing and stabilise at around $Q \sim 0.07$ Å^{-1} from 90 ns onwards. As discussed above the high Q liquid structure peak appears at around $Q \sim 1.5$ Å^{-1} due to the molecular structure factor peak observed in neutron diffraction measurements of neat D_2O. The difference in the low Q peak position with experimental data can be related to the micellar size and distribution observed as previously explained in the Section 4.2.2.

We can now understand the "intermediate Q" peak in the case of hybrid particle field simulations of SDS surfactants, as is seen developing in Fig. 17b at $Q \sim 0.6$ Å^{-1}, by analogy to the molecular structure factor peak for CTAB, noting that for the hybrid particle field simulations the potentials used are much more diffuse. This leads to a significant overlap of the beads and very little in terms of structure, as shown in the much

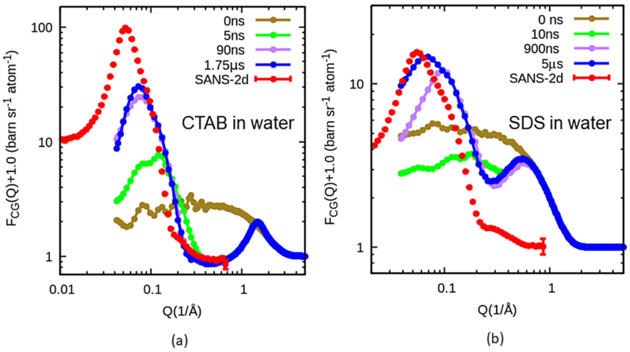


Fig. 17 (a) $F_{CG}(Q)$ plotted as the simulation progressed with time for (a) 100 mM of CTAB in water (b) 60 mM of SDS in water.

flatter water–water $g(r)$ (Fig. S20 in SI). At $t = 0$ both simulations show similar scattering patterns that quickly plateau, the only difference being the Q-scale at which this happens due to the different size of the water beads (4 waters in SDS case and 2 waters in CTAB case). This scattering is due to density fluctuations in the randomly placed beads in the box. In the DPD simulation for CTAB this plateau almost instantly disappears due to the local intermolecular structure factor of water beads (see structured $g(r)$ in Fig. S20 of SI). Conversely, in the case of SDS the combination of a larger water bead, and high bead overlap (as shown through an unstructured $g(r)$ that does not go to zero at low r), means this plateau largely remains, forming an intermediate Q shoulder on the lower Q micelle structure factor peak. Scattering calculated from pure water hpF-MD simulations and the presence of the shoulder between 0.3–1 Å^{-1} supports this argument (details are given in Section S10 of SI).

To further explore the reasons for the differences between the calculated scattering from CG simulations and the SANS experiments we return to the validation approach shown in Section 4.1, to confirm that differences between the simulation and experiment are due to the CG simulation methodology, rather than the calculation method. To achieve this, we performed necessarily short atomistic simulations of the systems and then generated pseudo-CG trajectories, following the methods used in the validation tests. The SASview analysis of the SANS-2d data has given an estimate of nearly 60 surfactants per micelle in 60 mM SDS case and 135 surfactants per micelle in 100 mM CTAB case. We therefore performed GROMACS simulations on an atomistic system with micelles of size suggested by SASview, built using the Shapespyer tool.[61] A single micelle surrounded by water has been generated and the structure has been well equilibrated for 100 ns using GROMACS. The single micelle was replicated in all directions generating a large system of 8 micelles in water. This was then equilibrated again for another 10 ns using GROMACS. Fig. 18 and 19 show the $F(Q)$ obtained from the atomistic trajectory of 8 preformed micelles and $F_{CG}(Q)$ from pseudo CG-trajectories compared with SANS-2d data for SDS and CTAB cases respectively.

Fig. 18 and 19 show better match with experiments when compared to CG simulations shown in Sections 4.2.1 and 4.2.2. Although ripples are present in the data due to the shorter simulation time, and therefore reduced exploration of the ensemble as necessitated by the more expensive atomistic calculation. The intermediate peak is absent in the case of SDS in water. This test confirms that the differences are due to the coarse-grained simulation methodology but not from the scattering calculation. The error in the low Q region shown in the Fig. 18 and 19 highlights the differences in micellar properties like aggregate size, shape and polydispersity. The calculation of scattering from the atomistic based pseudo-CG simulations also show how atomistic simulations are insufficient to predict scattering, due to their slow dynamics resulting in scattering data from an insufficiently large ensemble of structures.

The purpose of these comparison studies is to show the ability of MuSSIC to obtain such quantitative information on micellar properties of surfactant systems from CG simulations which allow direct comparison to experiments. This study not only assesses the accuracy of the CG potentials and parameters in the force fields from the differences with the experiments but also, potentially allows us to tune those potentials to match with experiments, allowing the determination of a data-refined structure,[36] and/or avenues for development of new CG force-fields based on experimental structural data.

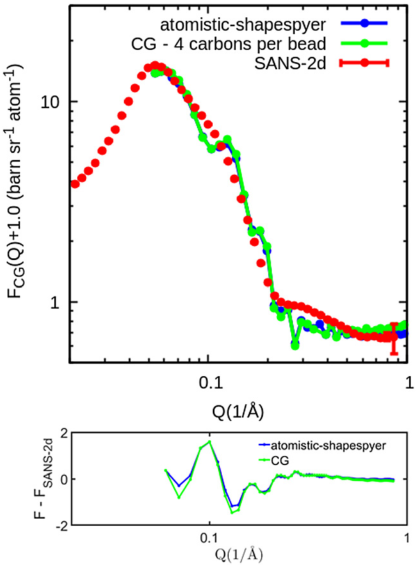


Fig. 18 $F(Q)$ obtained from the atomistic trajectory of preformed SDS micelles in water using shapespyer and equilibrated using GROMACS. $F_{CG}(Q)$ is computed from pseudo CG-trajectory and compared against SANS-2d data. The difference between the SANS-2d data and the simulation is shown in the bottom plot as an estimate of error.

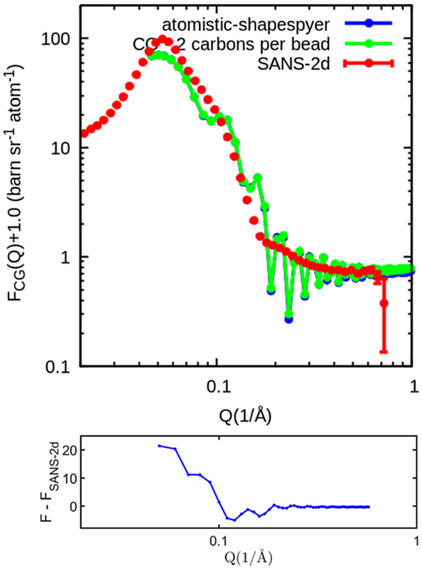


Fig. 19 $F(Q)$ and $F_{CG}(Q)$ obtained from the atomistic trajectory of preformed CTAB micelles in water using shapespyer are compared against SANS-2d data. The difference between the SANS-2d data and the simulation is shown in the bottom plot as an estimate of error.

5 Conclusions and outlook

MuSSIC is a computational tool for the calculation of neutron-weighted structure factors from CG molecular simulation systems, affording direct comparison of the simulations to a rigorous benchmark of structural experimental data. It has been designed specifically for high concentration soft matter systems where the structure of the solvent cannot be ignored. The code computes the partial pair radial distribution functions and the partial pair structure factors for the trajectory provided in *xyz* file format. The code has been verified using different CG models of surfactants in water and complex polymer melt systems, against atomistic data from the same underlying structure. In general, there is a close match to the atomistic calculation for $Q < 0.5$ Å^{-1}, with increasing divergence at higher Q with increasing bead size, as is expected by the loss in atomistic resolution. In particular it was observed that for the HD polymer example explored, that there was a notable sharp increase in disagreement for beads containing 6 heavy atoms, due to the CG structure losing it's ability to remain commensurate with the underlying atomtic structure. We have therefore demonstrated the validity of the method for calculating small-angle neutron scattering, on an absolute scale, for Q-ranges of SANS instruments, which typically have a maximum Q of around 0.5 Å^{-1}. We note that an alternative approach is to "backmap" the CG simulation to an atomistic one. While this would result in potentially higher accuracy in the scattering calculation, it would come at significant additional computational cost of at least a factor of 10 due to the increased number of scattering centers. Our validation tests show that in the typical SANS Q-range relative error between CG and atomistically calculated scattering for the polyamide case is less than 13% for $Q < 0.5$ Å^{-1} and below 1% at $Q = 0.1$ Å^{-1} (with much lower error for the C_{10}TAB case), tending to zero in relative and absolute terms as Q decreases. Nevertheless, a back mapping approach would be needed for accurate calculation of scattering for wide-angle scattering to replicate the nearest neighbour intermolecular liquid structures properly. We also note that a "forward-mapping" approach from atomistic to CG as used in the validation tests could be used as a way of calculating SANS data, by reducing the number of scattering centres and therefore reducing the complexity of the calculation.

The scientific insight afforded by use of the code has been demonstrated on large DPD and hPF-MD simulations of C_{16}TAB and SDS solutions respectively. In both cases, the structural differences between the simulation and experiment were clearly described. This level of information is vital in testing the appropriateness and transferability of CG force-fields, where results can be highly dependent on what data force-field parameters have been derived from. It is clear, from these two examples at least, that typical CG simulation force-fields may not provide the best possible representation of structure as measured by SANS. In this sense the MuSSIC code also provides a well-verified scattering calculation tool as a first step towards developments of CG refinement methods, that are designed to "push" a simulation towards matching the experimentally measured structure – this could be a CG version of EPSR/Dissolve, or other novel algorithms that may be better suited to the longer length scales involved *e.g.* conformation searching for larger molecules. Ultimately, for wide Q-range neutron scattering instruments such as NIMROD, refinement of a CG simulation against the lower Q data could be combined with the refinement of an atomistic simulation, that is kept structurally coherent with the CG simulation, providing a structure refinement across multiple lengthscales. In addition to refinement methods, future plans to further develop the MuSSIC code include improved definition of bead form factors, support for non-cubic simulation boxes, convolution of the result with a instrumental resolution function and calculation of X-ray scattering.

A parallel question is how can CG simulation help with SANS experiments and data analysis? In the first instance, CG models could be used to test if structural differences will be sufficient to provide different scattering signals, helping to plan experiments. Furthermore, if the experiment does match the simulated scattering (or can be made to match through some form of refinement) it can provide a more robust structural model than standard fitting-based approaches, as the result is known to be consistent with underlying molecular geometries and (assuming a reasonable force-field) some sensible description of the intermolecular forces. Finally, the use of a molecular scale model can potentially form the vital link between different experimental techniques (*e.g.* NMR cryo-EM, coherent diffraction imaging) to co-refine to a structure, this may become increasingly important as the complexity of the systems studied increases for example in determining structures such as lipid nanoparticles and vesicles.[62,63]

Author contributions

HBK: data curation, formal analysis, investigation, methodology, software, validation, writing – original draft. GJS: investigation, methodology, validation, writing – reivew and editing. JD: investigation, validation, writing – review and editing. TGAY: methodology, validation, writing – review and editing. TFH: conceptualisation, funding acquisition, methodology, project administration, supervision, validation, writing – review and editing.

Conflicts of interest

There are no conflicts to declare.

Data availability

All simulated scattering data were calculated using MuSSIC v1.1, available as a release on github (**https://github.com/disorderedmaterials/MuSSIC/releases/tag/v1.1**). The SANS-2d data shown for the use cases in this study are available in MuSSIC repository at [**https://github.com/disorderedmaterials/MuSSIC/tree/d9c14d0130a333ad204f7e9ff95b06feb82f00f7/usability_tests**], reference number.[37] The authors confirm that the links to the experimental data supporting the findings of this study are available within the SI.

A description of the scattering calculation broken down into intra- and inter-molecular parts, a detailed description of the bead form factors used, further details of the atomistic and CG simulations, further details on the effect of bead size on the scattering calculation, a presentation of the residuals and R-factor for the scattering calculation (Cg *vs.* atomistic), plots of the intra- and intermolecular radial distribution functions for the CG-polymer beads, plots of the average aggregation numbers over time for the CG simulations, and details of the hybrid particle field simulation of pure 4-bead water. See DOI: **https://doi.org/10.1039/d5cp01390a**

Acknowledgements

This work has been funded and supported by the STFC Ada Lovelace Centre. We thank Andrey Brukhno (STFC scientific computing) for providing the initial self-assembled structures of C_{10}TAB surfactants using the Shapespyer toolchain. We thank Paola Carbone for providing simulation files for polyamide-66. The SCARF computing cluster provided by STFC was used to perform molecular dynamics simulations.

Notes and references

1 F. Schmid, *ACS Polym. Au*, 2023, **3**, 28–58.
2 A. Gooneie, S. Schuschnigg and C. Holzer, *Polymers*, 2017, **9**, 16.
3 M. Laurati, M. Tassieri, G. Espinosa, M. S. Oliveira and P. Joseph, *Front. Phys.*, 2022, **10**, 1111099.
4 K. J. Edler and D. T. Bowron, *Curr. Opin. Colloid Interface Sci.*, 2015, **20**, 227–234.
5 A. H. Larsen, Y. Wang, S. Bottaro, S. Grudinin, L. Arleth and K. Lindorff-Larsen, *PLoS Comput. Biol.*, 2020, **16**, e1007870.
6 A. Arbe, F. Alvarez and J. Colmenero, *Polymers*, 2020, **12**, 3067.
7 S. J. Perkins, D. W. Wright, H. Zhang, E. H. Brookes, J. Chen, T. C. Irving, S. Krueger, D. J. Barlow, K. J. Edler, D. J. Scott, N. J. Terrill, S. M. King, P. D. Butler and J. E. Curtis, *J. Appl. Crystallogr.*, 2016, **49**, 1861–1875.
8 D. W. Wright, E. L. Elliston, G. K. Hui and S. J. Perkins, *Biophys. J.*, 2019, **117**, 2101–2119.
9 K. L. Sarachan, J. E. Curtis and S. Krueger, *J. Appl. Crystallogr.*, 2013, **46**, 1889–1893.
10 T. F. Headen and M. P. Hoepfner, *Energy Fuels*, 2019, **33**, 3787–3795.
11 H. M. Cezar and M. Cascella, *J. Chem. Inf. Model.*, 2023, **63**, 4979–4985.
12 A. Soper, *Chem. Phys.*, 1996, **202**, 295–306.
13 A. K. Soper, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2005, **72**, 104204.
14 A. Soper, *Empirical Potential Structure Refinement – EPSRshell: a user's guide*, **https://purl.org/net/epubs/work/56239**.
15 T. Youngs, *Mol. Phys.*, 2019, **117**, 3464–3477.
16 T. F. Headen, C. A. Howard, N. T. Skipper, M. A. Wilkinson, D. T. Bowron and A. K. Soper, *J. Am. Chem. Soc.*, 2010, **132**, 5735–5742.
17 A. K. Soper, *ISRN Phys. Chem.*, 2013, **2013**, 1–67.
18 C. Hardacre, J. D. Holbrey, S. E. J. McMath, D. T. Bowron and A. K. Soper, *J. Chem. Phys.*, 2002, **118**, 273–278.
19 O. S. Hammond, D. T. Bowron and K. J. Edler, *Green Chem.*, 2016, **18**, 2736–2744.
20 A. Bernasconi, M. Dapiaggi, A. Pavese, D. T. Bowron and S. Imberti, *J. Phys. Chem. B*, 2012, **116**, 13114–13123.
21 S. E. McLain, A. K. Soper, A. E. Terry and A. Watts, *J. Phys. Chem. B*, 2007, **111**, 4568–4580.
22 A. K. Soper and D. T. Bowron, *Chem. Phys. Lett.*, 2017, **683**, 529–535.
23 R. Hargreaves, D. T. Bowron and K. Edler, *J. Am. Chem. Soc.*, 2011, **133**, 16524–16536.
24 H. Laurent, M. D. G. Hughes, M. Walko, D. J. Brockwell, N. Mahmoudi, T. G. A. Youngs, T. F. Headen and L. Dougan, *Biomacromolecules*, 2023, 4869–4879.
25 D. T. Bowron, A. K. Soper, K. Jones, S. Ansell, S. Birch, J. Norris, L. Perrott, D. Riedel, N. J. Rhodes, S. R. Wakefield, A. Botti, M.-A. Ricci, F. Grazzi and M. Zoppi, *Rev. Sci. Instrum.*, 2010, **81**, 033905.
26 R. K. Heenan, S. M. King, D. S. Turner and J. R. Treadgold, *Proc. ICANS-XVII*, 2006, 780–785.
27 A. J. Smith, S. G. Alcock, L. S. Davidson, J. H. Emmins, J. C. H. Bardsley, P. Holloway, M. Malfois, A. R. Marshall, C. L. Pizzey, S. E. Rogers, O. Shebanova, T. Snow, J. P. Sutter, E. P. Williams and N. J. Terrill, *J. Synchrotron Radiat.*, 2021, **28**, 939–947.
28 J. E. Curtis, S. Raghunandan, H. Nanda and S. Krueger, *Comput. Phys. Commun.*, 2012, **183**, 382–389.
29 H. Iqbal, K. W. Fung, J. Gor, A. C. Bishop, G. I. Makhatadze, B. Brodsky and S. J. Perkins, *J. Biol. Chem.*, 2023, **299**, 102799–102800.
30 V. A. Spiteri, J. Doutch, R. P. Rambo, J. Gor, P. A. Dalby and S. J. Perkins, *Biophys. J.*, 2021, **120**, 1814–1834.
31 L. O. Puster, C. B. Stanley, V. N. Uversky, J. E. Curtis, S. Krueger, Y. Chu and C. B. Peterson, *Biochemistry*, 2019, **58**, 5117–5134.
32 A. Arbe, F. Alvarez and J. Colmenero, *Soft Matter*, 2012, **8**, 8257–8270.
33 M. G. Guenza, M. Dinpajooh, J. McCarty and I. Y. Lyubimov, *J. Phys. Chem. B*, 2018, **122**, 10257–10278.
34 S. Takada, *Curr. Opin. Struct. Biol.*, 2012, **22**, 130–137.
35 F. R. Souza, L. M. P. Souza and A. S. Pimentel, *J. Chem. Inf. Model.*, 2020, **60**, 5881–5884.
36 A. K. Soper and K. J. Edler, *Biochim. Biophys. Acta*, 2017, **1861**, 1652–1660.
37 H. B. Kolli, T. F. Headen, G. Jimenez-Serratos and T. Youngs, *MuSSIC: software for neutron Scattering calculation for coarse grain models of soft matter*, 2023, **https://github.com/disorderedmaterials/MuSSIC**.
38 D. A. Keen, *J. Appl. Crystallogr.*, 2001, **34**, 172–177.
39 G. Milano and T. Kawakatsu, *J. Chem. Phys.*, 2009, **130**, 214106.

40 K. Schäfer, H. B. Kolli, M. Killingmoe Christensen, S. L. Bore, G. Diezemann, J. Gauss, G. Milano, R. Lund and M. Cascella, *Angew. Chem., Int. Ed.*, 2020, **59**, 18591–18598.

41 A. De Nicola, Y. Zhao, T. Kawakatsu, D. Roccatano and G. Milano, *J. Chem. Theory Comput.*, 2011, **7**, 2947–2962.

42 G. J. A. Sevink, F. Schmid, T. Kawakatsu and G. Milano, *Soft Matter*, 2017, **13**, 1594–1623.

43 S. L. Bore, H. B. Kolli, A. De Nicola, M. Byshkin, T. Kawakatsu, G. Milano and M. Cascella, *J. Chem. Phys.*, 2020, **152**, 184908.

44 S. L. Bore, H. B. Kolli, T. Kawakatsu, G. Milano and M. Cascella, *J. Chem. Theory Comput.*, 2019, **15**, 2033–2041.

45 Y.-L. Zhu, Z.-Y. Lu, G. Milano, A.-C. Shi and Z.-Y. Sun, *Phys. Chem. Chem. Phys.*, 2016, **18**, 9799–9808.

46 Y. Zhao, A. De Nicola, T. Kawakatsu and G. Milano, *J. Comput. Chem.*, 2012, **33**, 868–880.

47 P. Hoogerbrugge and J. Koelman, *EPL*, 1992, **19**, 155.

48 R. D. Groot and P. B. Warren, *J. Chem. Phys.*, 1997, **107**, 4423–4435.

49 P. Espanol and P. B. Warren, *J. Chem. Phys.*, 2017, **146**, 150901.

50 A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott and S. J. Plimpton, *Comput. Phys. Commun.*, 2022, **271**, 108171.

51 M. Svoboda, M. G. Jiménez, A. Kowalski, M. Cooke, C. Mendoza and M. Lsal, *Soft Matter*, 2021, **17**, 9967–9984.

52 R. L. Anderson, D. J. Bray, A. S. Ferrante, M. G. Noro, I. P. Stott and P. B. Warren, *J. Chem. Phys.*, 2017, **147**, 094503.

53 R. L. Anderson, D. J. Bray, A. Del Regno, M. A. Seaton, A. S. Ferrante and P. B. Warren, *J. Chem. Theory Comput.*, 2018, **14**, 2633–2643.

54 O. Arnold, J. C. Bilheux, J. M. Borreguero, A. Buts, S. I. Campbell, L. Chapon, M. Doucet, N. Draper, R. Ferraz Leal, M. A. Gigg, V. E. Lynch, A. Markvardsen, D. J. Mikkelson, R. L. Mikkelson, R. Miller, K. Palmen, P. Parker, G. Passos, T. G. Perring, P. F. Peterson, S. Ren, M. A. Reuter, A. T. Savici, J. W. Taylor, R. J. Taylor, R. Tochenov, W. Zhou and J. Zikovsky, *Nucl. Instrum. Methods Phys. Res., Sect. A*, 2014, **764**, 156–166.

55 *Mantid-Data analysis software*, https://www.mantidproject.org/installation/index.

56 SasView for Small Angle Scattering Analysis, https://www.sasview.org.

57 J. B. Hayter and J. Penfold, *Mol. Phys.*, 1981, **42**, 109–118.

58 J.-P. Hansen and J. B. Hayter, *Mol. Phys.*, 1982, **46**, 651–656.

59 H. B. Kolli, A. de Nicola, S. L. Bore, K. Schäfer, G. Diezemann, J. Gauss, T. Kawakatsu, Z.-Y. Lu, Y.-L. Zhu, G. Milano and M. Cascella, *J. Chem. Theory Comput.*, 2018, **14**, 4928–4937.

60 R. Mao, M.-T. Lee, A. Vishnyakov and A. V. Neimark, *J. Phys. Chem. B*, 2015, **119**, 11673–11683.

61 J. D. Andrey Brukhno, *Shapespyer: Python-driven toolchain for modelling soft matter*, 2023, https://gitlab.com/simnavi/shapespyer.

62 D. L. Pink, O. Loruthai, R. M. Ziolek, P. Wasutrasawat, A. E. Terry, M. J. Lawrence and C. D. Lorenz, *Small*, 2019, **15**, 1903156.

63 V. Nele, M. N. Holme, M. H. Rashid, H. M. Barriga, T. C. Le, M. R. Thomas, J. J. Doutch, I. Yarovsky and M. M. Stevens, *Langmuir*, 2021, **37**, 11909–11921.