Soft Matter

RSC Publishing

PAPER

View Article Online
View Journal | View Issue

How to calculate structure factors of self-assembling anisotropic particles

Cite this: *Soft Matter*, 2013, **9**, 4412

Sofia Kantorovich,*[abc] Elena Pyanzina,[a] Cristiano De Michele[b] and Francesco Sciortino[b]

We put forward a theoretical approach to analyse the structure factors obtained experimentally for solutions of self-assembling anisotropic particles. This method is applicable for any system of particles forming chains in thermodynamic equilibrium and is based on studying the behaviour of the centre–centre structure factor first peak. In order to calculate the structure factor analytically, we first derive the radial distribution function. For that we use the combination of the density functional theory and density expansion of the pair correlation functions. The first theory allows for the equilibrium chain formation, and the second takes into account particles' shape anisotropy at the level of the Gay–Berne potential. We apply our method to describe centre–centre structure factors of self-assembling short DNA duplexes with various semiaxes ratios. Extensive comparison of the theoretical predictions with the Monte Carlo simulation data shows a very good agreement. We show that the particle shape anisotropy exerts a crucial influence on the behaviour of the structure factors.

Received 18th December 2012
Accepted 8th February 2013

DOI: 10.1039/c3sm27895f

www.rsc.org/softmatter

1 Introduction

The spontaneous thermodynamic equilibrium formation of reversible aggregates from basic building blocks, whose size may range from a few angstroms to microns, is often referred to as self-assembly. Since molecules, macromolecules or colloidal particles are all possible examples of building blocks, self-assembly is ubiquitous in nature and of interest in many fields of research, including materials science, soft matter and biophysics. Self-assembly is one of the most promising routes for the realisation of novel materials, because tuning the shape, valence, flexibility and mutual interactions of the individual building blocks can finely control the physical properties of these new bottom-up designed materials. In addition, the material properties can be tailored for replicating on different time- and length-scales the behaviour of atomic systems. In fact, one of the dreams of soft matter scientists would be to use such a wide portfolio of physical and chemical interactions to synthesise materials with novel properties and functions, the smart materials of the future. Of particular interest is the situation in which the anisotropy of the attractive potential may favour the formation of linear aggregates. The mutual alignment of these chains, if they possess sufficient rigidity, might lead to the formation of macroscopically orientationally ordered liquid crystal phases (*e.g.* nematic or columnar) above a critical concentration.

Relevant examples of self-assembling building blocks are the approximately cylindrically shaped DNA duplexes that form as a result of hybridisation of short complementary sequences.[1] These short DNA duplexes, due to coaxial stacking interactions between their blunt ends, self-assemble into weakly bonded linear chains.[1,2] This behaviour is not restricted to B-form DNA oligomers, as it has also been observed in solutions of blunt-ended A-form RNA oligomeric duplexes. Analogous cases are those of chromonic liquid crystals,[3–5] worm-like micelles of amphiphilic molecules in water or micro-emulsions of water and oil which are stabilised by amphiphilic molecules,[6–8] fibres and fibrils of peptides and proteins,[9–11] and, finally, the G-quadruplexes forming very rigid cylindrical chains.[12,13] A deeper physical understanding of these stacking interactions is very important for different reasons. Self-assembly is also invoked as a prebiotic route[14] for the formation of complex molecules, possibly RNA oligomers, from elementary carbon-based molecules. Besides that, the self-assembly of DNA duplexes provides a suitable way to access and quantify hydrophobic coaxial stacking interactions. Finally, the LC ordering of nucleic acids can be viewed as a new model of reversible aggregation leading to a macroscopic ordering, where the phase diagram of the system can be altered by choosing a different sequence for the DNA oligomers and/or introducing terminal overhangs in the duplexes.

The range of potential applications of self-assembling of particles, molecules, macromolecules, *etc.* is very broad. Thus, for example, G-quadruplexes can serve in medicine as anti-cancer drugs because of their ability to inhibit telomerase.[15–18] Chromonic mesophases at high concentrations can be found in aqueous dispersions of many formulated products, such as

[a]*Ural Federal University, Lenin av. 51, Ekaterinburg, Russian Federation. E-mail: sofia.kantorovich@univie.ac.at*
[b]*University of Rome "La Sapienza", P.le A. Moro 2, Roma, Italy*
[c]*University of Vienna, Sensengasse 8, Wien, Austria*

pharmaceuticals and dyes used in inkjet printing. In the last years, chromonic materials attracted a lot of attention due to their potential use as functional materials for fabricating highly ordered thin films[19] and as biosensors.[20] In addition, self-assembly of amyloid protein fibrils is a mechanism responsible for many human diseases, like Alzheimer's disease, Huntingtons disease, Creutzfeld–Jakob disease, Parkinson's disease and type 2 diabetes.[21]

Recently, we investigated the reversible physical polymerisation and collective ordering of DNA duplexes by modelling them as quasi-cylindrical superquadrics[22–24] with two reactive sites on their bases. Therein, we presented a theoretical framework which turned out to be able to properly account for their reversible association process. In the subsequent work,[23] we studied a more realistic model of DNA, proving the validity of the developed method. Based on the aforementioned studies, here we provide a theoretical strategy to calculate the static structure factor and radial distribution functions for systems which self-assemble into linear aggregates through a physical polymerisation. This method is very affordable compared to much more complicated alternatives like integral equation theories[25–27] or full atomistic[28,29] or even coarse-grained simulations,[30,31] for instance. It is based on the direct calculation of the pair correlation function for a system, for which the equilibrium chain distribution is known to be exponential.[32–34] Additionally, we propose a way to analytically include particle shape anisotropy in the pair correlation function. Our method with its bottom-up strategy might provide the desired framework, in which the relationship between the macroscopic response of the system and its microscopic structure might be accurately interpreted. In this respect, it can be of relevance for interpreting data extracted from scattering patterns as well as from rheological or birefringence measurements. More specifically, we focus on the spacial region corresponding to nearest neighbour distances (and its corresponding wave vector range), rather than large distances (*i.e.* small wave vector).

Although direct comparison with the experimental data is not possible due to the lack of direct experimental measurements for the chain-forming anisotropic building blocks, experiments using Small Angle Neutron Scattering (SANS) to elucidate various self-assembling systems are already available.[35,36] On the other hand, future comparison between scattering experiments and our approach might provide a method to estimate the average polymerisation and the associated cluster size distributions. We indeed show that from the structure factor curves, for a system with known density and building-block shape one can uniquely determine the average chain length and cluster size distribution. Note that the unique interpretation becomes possible only if the system under study exhibits reversible self-assembly in linear chain structures in the thermodynamic equilibrium. If the topology of clusters is not known *a priori*, or the system is out of equilibrium, the structure factor, being an integral property, can be rather misleadingly interpreted.

This manuscript is organised as follows. First we provide an overview of the density functional theory for self-assembling short DNA duplexes. In this overview we mainly focus on the main results, rather than explain the way to obtain them. For details one can address ref. 22 and references therein. In Section 3 we put forward an approach to calculate the pair correlation function (radial distribution function, RDF) for anisotropic particles, which can form linear chains. Here, we also extensively compare theoretical predictions to the results of Monte Carlo simulations, and analyse the influence of the shape anisotropy on the RDFs. Based on the previous analytically obtained RDFs, in Section 4 we calculate the structure factor for systems of short DNA duplexes at different temperatures, densities and values of shape anisotropy. In the second part of this section we carefully analyse the relationship between the microstructure characteristics of the systems and their structure factors. The summary of the work is given in the Conclusion. At the end we provide Appendix 1 and Appendix 2 with all important details concerning analytical calculations of the radial distribution function for a system of anisotropic particles. This manuscript is focused on the isotropic phase, meaning that the model developed here is valid only for relatively low densities at which the chains of short DNA duplexes are forming a dilute fluid phase. The other region of the phase diagram, where the transition to the nematic ordering might occur will be the subject of a separate study.

2 What is known theoretically about self-assembly of DNA-duplexes in the isotropic phase?

In this part of the manuscript we summarise previous results of Density Functional Theory (DFT) and Monte Carlo (MC) simulations[22] in order to characterise the microscopic structure, namely the chain size distribution and the average chain length. DNA duplexes are approximated by superquadrics (SQs), *i.e.*, generalised hard ellipsoids (HEs) whose surface is defined by

$$f(x,y,z) = \left|\frac{x}{a}\right|^p + \left|\frac{y}{b}\right|^m + \left|\frac{z}{c}\right|^n - 1 = 0, \quad (1)$$

where the parameters p, m, n are real numbers and a, b, c are the SQ semiaxes. We set $m = n = 2$, $p = 16$ and $b = c$, so that the SQ resembles a cylinder with rounded edges (see ref. 22 for further details). The aspect ratio of a SQ is $X_0 = a/b$ and, as for the case of HEs, SQs of aspect ratio $X_0 < 1$ are called "oblate", while SQs of aspect ratio $X_0 > 1$ are called "prolate". In the present study we investigated only prolate SQs with aspect ratio $X_0 = 1$, 2 and 3. Furthermore each particle is decorated with two attractive sites, located along the symmetry axis at a distance $(X_0 - 0.46)b$ from the centre of mass, so that the anisotropic hard-core interaction is complemented with an anisotropic attractive potential, as previously done for modelling water,[37] silica[38] and the stepwise polymerisation of bifunctional diglycidyl ether of bisphenol A with pentafunctional diethylenetriamine.[39,40]

For this model, earlier we developed a theoretical approach and performed MC simulations to analyse the microscopic structure and self-assembly for various temperatures, or equivalently different interaction strengths, and densities of SQs.[22]

Our simulations were performed in the NVT ensemble. As a unit of length in our simulations we used the radius of the SQ main cross-section (the diameter of the SQ was $D = 2$ nm.) To detect the overlap of two SQs we calculated the distance using the algorithm described in ref. 24. We studied a system of $N = 1000$ particles in a wide range of volume fractions ϕ and temperatures. In all simulations we adopted periodic boundary conditions in a cubic simulation box. From the stored equilibrium configuration we have calculated the RDF and structure factor $S(q)$ to be compared with the prediction of the present manuscript (see next sections).

Based on the early work of ref. 6, 41 and 42 and the more recent developments,[22,23] the free energy F of a system of equilibrium polymers in the isotropic phase can be written as

$$\frac{\beta F}{V} = \sum_{l=1}^{\infty} \nu(l,\phi,T,X_0)\{\ln[v_d \nu(l,\phi,T,X_0)] - 1\} + \frac{\eta(\phi)}{2} \sum_{\substack{l=1 \\ l'=1}}^{\infty} \nu(l,\phi,T,X_0)\nu(l',\phi,T,X_0)\, v_{excl}(l,l',X_0) - \beta\Delta F_b \sum_{l=1}^{\infty}(l-1)\nu(l,\phi,T,X_0), \quad (2)$$

where V is the volume of the system, $\beta = 1/k_B T$, with k_B denoting the Boltzmann constant, and T being the temperature, v_d is the volume of a monomer, $\phi \equiv v_d\rho$ ($\rho = N/V$ is the number density of monomers, N is the number of monomers) is the packing fraction. The function $\nu(l, \phi, T, X_0)$ is the number density of chains of length l and it is assumed to be exponential, *i.e.*

$$\nu(l, \phi, T, X_0) = \rho M^{-(l+1)}(M-1)^{l-1}, \quad (3)$$

where M is the average chain length. The parameter ΔF_b is the free energy associated with the formation of a bond, and for the present SQ model it has the form:

$$\beta\Delta F_b = \beta\Delta E_S + \sigma_b. \quad (4)$$

Here, ΔE_S is the bonding energy and σ_b is the bonding entropy. In eqn (2) $v_{excl}(l, l', X_0)$ is the excluded volume of two chains of length l and l', *i.e.*

$$v_{excl}(l,l',X_0) = 2B_I X_0^2 ll' + 2v_d k_I \frac{l+l'}{2}, \quad (5)$$

where the parameters $B_I(X_0)$ and $k_I(X_0)$ can be estimated *via* MC integrals of a system composed of only two monomers (see ref. 22). $\eta(\phi)$ is the Parsons–Lee factor,[22] *i.e.*

$$\eta(\phi) = \frac{1}{4}\frac{4-3\phi}{(1-\phi)^2}. \quad (6)$$

The free energy functional eqn (2) explicitly accounts for the polydispersity inherent to the equilibrium polymerisation process using a discrete chain length distribution and for the entropic and energetic contributions of each single bond through the parameter ΔF_b.

With the aforementioned assumption for $\nu(l)$, the free energy in eqn (2) can be written as:

$$\frac{\beta F}{V} = -\rho\beta\Delta F_b(1 - M^{-1}) + \eta(\phi)\left[B_I X_0^2 + \frac{v_d k_I}{M}\right]\rho^2 + \frac{\rho}{M}\left[\ln\left(\frac{v_d\rho}{M}\right) - 1\right] + \rho\frac{M-1}{M}\ln(M-1) - \rho\ln M. \quad (7)$$

Minimisation of the free energy in eqn (7) with respect to M provides the following expression for the average chain length:

$$M \equiv M(\phi,T,X_0) = \frac{1}{2}\left(1 + \sqrt{1 + 4\phi e^{k_I\phi\eta(\phi)+\beta\Delta F_b}}\right). \quad (8)$$

Thus, we can analytically calculate the average chain length $M(\phi, T, X_0)$ and the chain size distribution $\nu(l, \phi, T, X_0)$ (see, eqn (3)). Earlier, theoretical predictions were compared to the simulation results and were found to be in a very good agreement.[22] So, below in this manuscript we use analytical expressions to characterise the system structure. For different semiaxes ratios, we summarise the influence of temperature and density on $M(\phi, T, X_0)$ and $\nu(l, \phi, T, X_0)$ by plotting colour (ϕ–T)-diagrams in Fig. 1. One can see that in the shown isotropic phases, the highest M value does not exceed four, even for the lowest T and the highest ϕ. M grows monotonously with $1/T$ and ϕ. The size of the chains only weakly, but nonmonotonously, depends on X_0. Thus, we observe the highest M values for

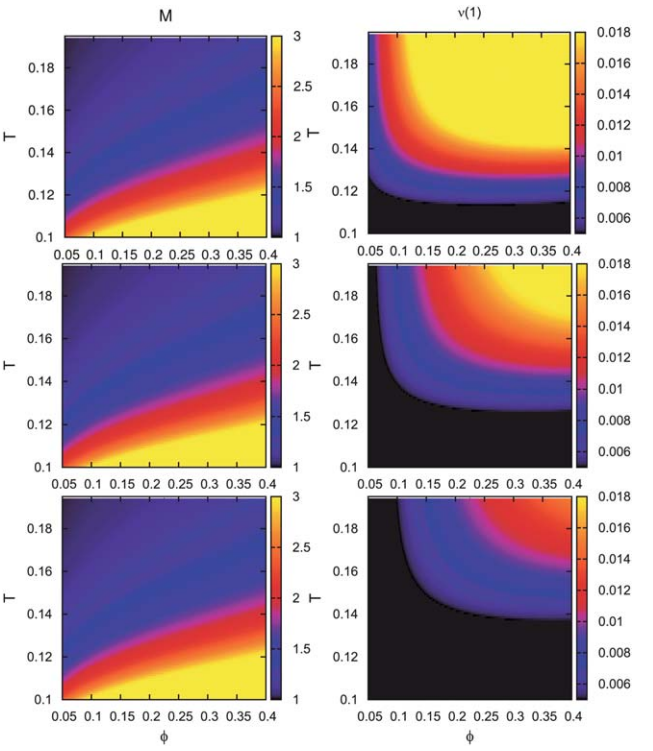

Fig. 1 Colour (ϕ–T)-diagrams characterising the microscopic structure of the SQ systems for three different semiaxes ratios: the first row corresponds to $X_0 = 1$; the second row is calculated for $X_0 = 2$; and the results for $X_0 = 3$ are provided in the third row. Left column: the colour corresponds to the average number of SQs in the chain $M(\phi, T, X_0)$ for the given temperature and density; the colour-code is shown on the right of each diagram. Right column: the colour corresponds to the number of nonaggregated SQs $\nu(1) \equiv \nu(1, \phi, T, X_0)$ for the given temperature and density; the colour-code is shown on the right of each diagram. For both columns the lighter the colour is, the higher is the value of the observable.

$X_0 = 1$; for $X_0 = 2$, M becomes smaller but starts growing again with increasing semiaxes ratio. The behaviour of the number of monomers $\nu(1, \phi, T, X_0)$ (single SQs) turns out to be more sophisticated. Obviously, independent of the shape anisotropy, the largest quantity of nonaggregated SQs can be found at the highest T. For almost all T, the higher ϕ is, the larger is $\nu(1, \phi, T, X_0)$. However, the ϕ dependence for a certain range of T becomes nonmonotonous: see, for example, the region of $0.12 < T < 0.14$. The total number of monomers is found to be a monotonously decreasing function of X_0. To summarise, for all three X_0 values we observe a reversible moderate chain formation, the chains are rather short for any particle anisotropy, and the amount of monomers is rather sensitive to all three possible control parameters (T, ϕ and X_0).

3 Isotropic approximation of anisotropic particles

Experimental studies of DNA solutions usually describe macroscopic properties of the system. In other words, if one is interested in the microscopic structure of these solutions, experimental data provide only explicit information, which needs to be interpreted employing a suitable theoretical model. One of the most widely used experimental tools to investigate microscopic details of the DNA solutions is small angle scattering. In this technique, after a suitable treatment of the data and subtraction of the form factor, one measures the structure factor $S(\boldsymbol{q})$, which is the Fourier transform of the centre-to-centre radial distribution function. Thus, it would be highly useful to develop a model describing radial distribution functions and structure factors of DNA duplex solutions on the basis of microscopic theory and simulations, where the structure of the solution could be directly accessed.

As we have already mentioned above, DNA duplexes form chains. Earlier, for the system of spherical dipolar chain-forming particles, we developed a model based on the fine weave of the DFT and molecular dynamics simulations which allowed for chain formation in thermodynamic equilibrium and revealed the connection between chain distribution and system structure factor.[33] This model is based on the assumption that the magnetic fluid is rather dilute, so that the main contribution to the pair correlation function is given by intra-chain correlations, rather than inter-chain interactions. Knowing the chain size distribution in thermodynamic equilibrium (using the DFT approach), one can directly calculate the number of particles contributing to the RDF, considering only distances accessible for chains, such as one diameter, two diameters, three diameters, *etc.* To obtain the structure factor one needs then just to Fourier transform a theoretically constructed RDF.

The aforementioned method proved to be very efficient to describe structure factors of mono- and bidisperse magnetic fluids with chain aggregates.[32–34] Exploiting the similarities in the microscopic structure, we decided to use the same approach and, on the basis of the theoretically evaluated $\nu(l, \phi, T, X_0)$ (described in the previous section), to calculate radial distribution functions for the systems of short DNA duplexes. We then compare the theoretical predictions with the "exact" RDFs and structure factors resulting from the analysis of MC configurations.

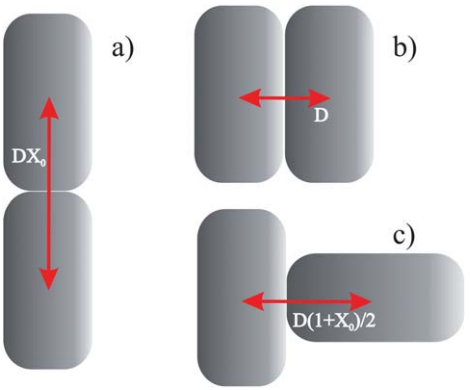

Fig. 2 The schematic drawing of SQs in head-to-tail, side-by-side and "T"-configurations. Corresponding centre–centre distances are given in the figure.

Looking at the radial distribution functions for different values of T and ϕ for all three semiaxes ratios investigated in the simulations (see, Fig. 3–5, symbols), however, one can immediately see the influence of the duplex anisotropy, a feature which was missing in the case of dipolar spherical particles.

A straightforward extension of the simple theory developed for spherical dipolar particles predicts that the pair correlation function is zero up to the distance approximately equal to one of the close-contact head-to-tail configurations (see Fig. 3–5, dashed lines and Fig. 2a). This is because in the approach proposed by Pyanzina and coauthors,[32–34] only the contributions coming from the intra-chain distances are allowed for. However, for the superquadrics, only for $X_0 = 1$, the first distance in the chain is equal to the first distance accessible for the particles (see Fig. 2). As it is schematically shown in Fig. 2, for $X_0 \neq 1$ the centre to centre distance can be smaller than the one in chains. Thus, for $X_0 \neq 1$, the first coordination shell corresponds to the side-by-side parallel configuration of the cylinders Fig. 2b. This configuration is highly entropically disadvantageous, but still is present in the systems with even slightly anisotropic duplexes. Another "T"-shape configuration (Fig. 2c), which corresponds to the distance of $D(X_0 + 1)/2$, on the contrary, is entropically rather advantageous. That is why with growing anisotropy ratio, from the first set of figures to the last, one can see that the RDF observed in simulations differs more and more significantly from the theoretical predictions based on the intra-chain distances only.

The main conclusion from the aforementioned set of figures is the following: the higher the value of X_0 is, the more pronounced becomes the shoulder in the RDF at the distances corresponding to the side-by-side and perpendicular "T"-configurations at close contact (see Fig. 2). For larger distances, however, the influence of anisotropy is not that strong, as the structure of the RDF is determined by chain formation, in particular, for low temperature. Comparing the T dependence of $\nu(1, \phi, T, X_0)$ from Fig. 1 with the T dependence of the RDF shoulder height in Fig. 3–5, one concludes that indeed the height of the shoulder in RDFs is directly correlated with

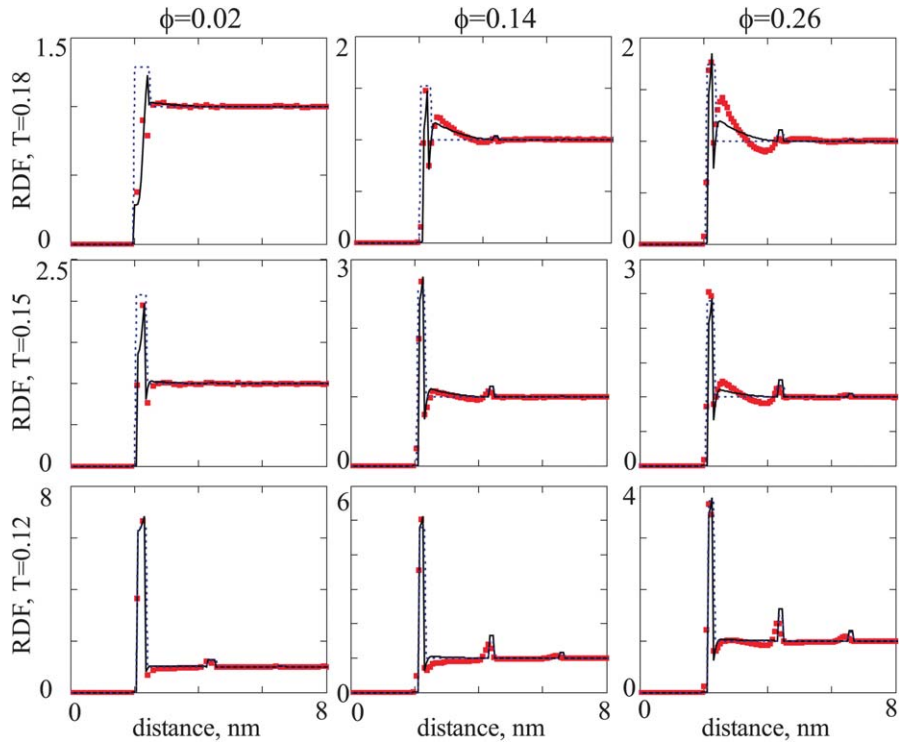

Fig. 3 Radial distribution functions for DNA self-assembling duplexes: $X_0 = 1$, symbols correspond to the simulation data; black solid lines describe the RDFs obtained theoretically using the approximated Gay–Berne potential together with the chain model; blue dashed lines are theoretical predictions, in which only chains are considered. The three rows correspond to three different temperatures: from the top to the bottom $T = 0.18$, $T = 0.15$ and, $T = 0.12$. Each column corresponds to the fixed density: from the left to the right $\phi = 0.02$, $\phi = 0.14$, $\phi = 0.26$. The diameter of the SQ is fixed to 2 nm.

monomer densities. Fig. 3–5 clearly evidence that the direct application of the approach developed in ref. 32–34 is impossible, due to the anisotropy of DNA duplexes, and it needs some adjustments at small distances.

One of the possible ways to take the anisotropy into account is to calculate the RDF of anisotropic particles at low distances, using its expansion in density.[43,44] The general form of the RDF expanded up to the first order in density ϕ is

$$g(r, T, X_0, \phi) = \beta_2(r, T, X_0) + \phi\beta_3(r, T, X_0). \quad (9)$$

Here the modulus of the distance $r = |\boldsymbol{r}|$ and $\beta_2(\,)$ and $\beta_3(\,)$ are coefficients analogous to the virial ones. For an arbitrary potential $U(\Omega, \boldsymbol{r})$, which depends on the interparticle distance $\boldsymbol{r}$ and mutual orientation of particles Ω, one can write $\beta_2(\,)$ as:

$$\beta_2(r, T, X_0) = \int \mathrm{d}\Omega \mathrm{d}\boldsymbol{r}^* \exp\left(-\frac{U(\Omega, \boldsymbol{r})}{k_\mathrm{B}T}\right) = \langle f(\Omega,\ \boldsymbol{r}) + 1\rangle_{12}.$$

In this expression, $\mathrm{d}\boldsymbol{r}^*$ is the angular part of the radius vector, connecting centres of two objects. In other words the averaging (denoted by angular brackets) of the Mayer function $f(\,) = \exp\left(-\frac{U(\Omega, \boldsymbol{r})}{k_\mathrm{B}T}\right) - 1$, is carried over all possible orientations, and angular degrees of freedom of the second particle, whereas the modulus of the distance remains a variable. The coefficient $\beta_3(r, T, X_0)$ in eqn (9) has the form:

$$\beta_3(r, T, X_0) = \int \mathrm{d}r_3 \langle (f(12) + 1) f(13) f(23)\rangle_{123}$$

The potential $U(\Omega, \boldsymbol{r})$ is determined by the problem itself. In the present study, we need to take into account particles' anisotropy. However, the SQ steric potential used in the simulations has no close functional form and as a consequence cannot be used to calculate the RDF analytically. To proceed we decided to select a potential which on the one hand retains the SQ shape but on the other hand can be written in a simple functional form. One of the most widely used potentials to mimic short-range interaction of Lennard-Jones type is the Gay–Berne potential.[45,46] It describes the interaction of two "soft" ellipsoidal particles and depends not only on the distance between their centres but also on their mutual orientation. The general form of the Gay–Berne potential is provided in Appendix 1. Its analytical form is rather cumbersome, but can be interpreted as a Weeks–Chandler–Andersen[47] potential with generalised interparticle distance. We used this fact to simplify the potential by introducing a single parameter s to replace the angular dependence of the interparticle distance. However, one needs to choose the parametrisation with care. In the present study we used $\beta_2(r, T, X_0)$ as a target function. In other words the "real value" of $\beta_2(r, T, X_0)$ was calculated as a function of r for various values of X_0 using the full anisotropic angular-dependent Gay–Berne potential. After that, polynomial parameterisation was found in such a way that the value of $\beta_2(r, T, X_0)$ for

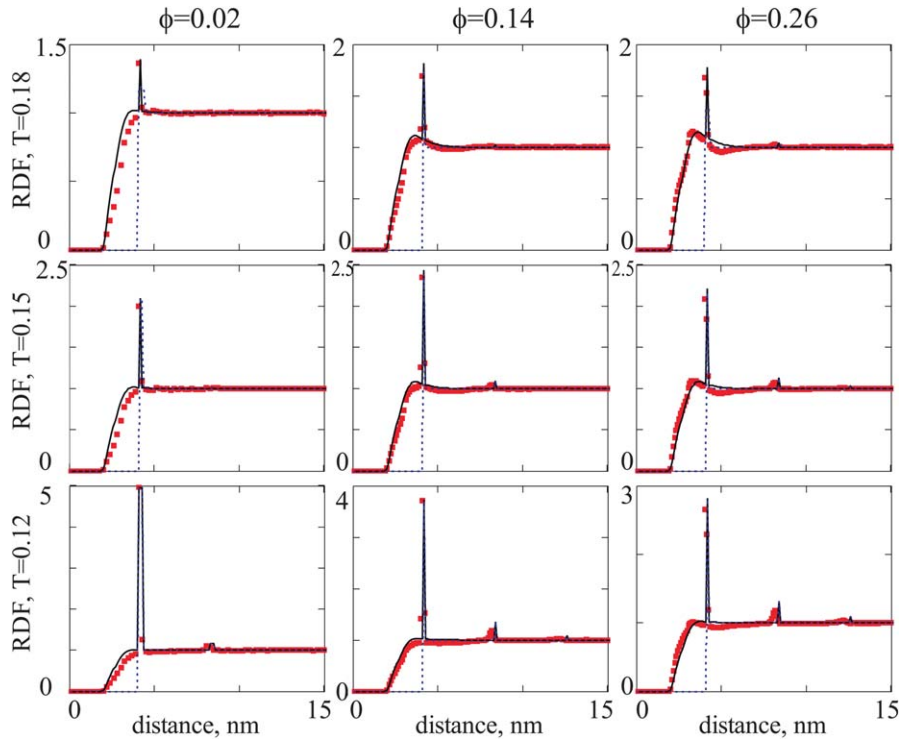

Fig. 4 Radial distribution functions for DNA self-assembling duplexes: $X_0 = 2$, symbols correspond to the simulation data; black solid lines describe the RDFs obtained theoretically using the approximated Gay–Berne potential together with the chain model; blue dashed lines are theoretical predictions, in which only chains are considered. The three rows correspond to three different temperatures: from the top to the bottom $T = 0.18$, $T = 0.15$ and $T = 0.12$. Each column corresponds to the fixed density: from the left to the right $\phi = 0.02$, $\phi = 0.14$, $\phi = 0.26$. The diameter of the SQ is fixed to 2 nm.

the simplified isotropic steric potential deviated from the original one by less than 2%. This approach allowed us to perform analytical calculations for $\beta_3(r, T, X_0)$, reducing the dimensionality of the integrals from 12 to three (for details, see Appendix 1). We assume that only nonaggregated SQs contribute to the previously discussed shoulder of RDFs. In this way, the resulting RDF becomes a superposition of three contributions: the contribution from free monomers, the contribution from the chain and the cross-contribution from the chain–monomer interaction. For the chains we use the following approximation:

$$g_{\mathrm{CH}}(r, T, X_0, \phi, \delta) = \frac{(\phi - \nu(1, \phi, T, X_0))^2}{\phi^2} \begin{cases} \beta_2(r, T, X_0), & \text{if } 0 < r < X_0 D - \delta; \\ \cdots & \\ \beta_2(r, T, X_0) + p^s / A_s(X_0, \phi, \delta), & \text{if } sX_0 D - \delta \le r \le sX_0 D + \delta; \\ \beta_2(r, T, X_0), & \text{if } sX_0 D + \delta < r < (s+1)X_0 D - \delta; \\ \cdots & \end{cases} \quad (10)$$

where $A_s(X_0, \phi, \delta) = 4\pi\phi(X_0 D)^2 s^2 \delta / v_{\mathrm{d}}$ and $p(T, X_0, \phi) = (M(T, X_0, \phi) - 1)/M(T, X_0, \phi)$. For monomer–monomer contribution the RDF can be written as:

$$g_{\mathrm{MM}}(r, T, X_0, \phi, \delta) = \frac{\nu(1, \phi, T, X_0)^2}{\phi^2} \times (\beta_2(r, T, X_0) + \nu(1, \phi, T, X_0)\beta_3(r, T, X_0)). \quad (11)$$

Finally, for cross-terms we make an assumption that in a three-body term, one SQ is always in a chain and two others are free monomers (this assumption stems from the fact that we are neglecting chain–chain correlations):

$$g_{\mathrm{MCH}}(r, T, X_0, \phi) = \frac{\nu(1, \phi, T, X_0)(\phi - \nu(1, \phi, T, X_0))}{\phi^2} \times (\beta_2(r, T, X_0) + \nu(1, \phi, T, X_0)\beta_3(r, T, X_0)). \quad (12)$$

Thus, the total RDF with particle anisotropy taken into account has the form (for details, see Appendix 2):

$$g(r, T, X_0, \phi, \delta) = g_{\mathrm{CH}} + g_{\mathrm{MM}} + 2g_{\mathrm{MCH}}. \quad (13)$$

In Fig. 3–5 we plot the RDF from eqn (13) with black solid lines. Theoretical predictions are now in very good agreement with the simulation data for a broad range of T and ϕ. One, however, should note that for $\phi = 0.26$ and $T = 0.18$, in the initial part, theoretical RDF deviates substantially from the simulation data and does not describe the shape of the shoulder

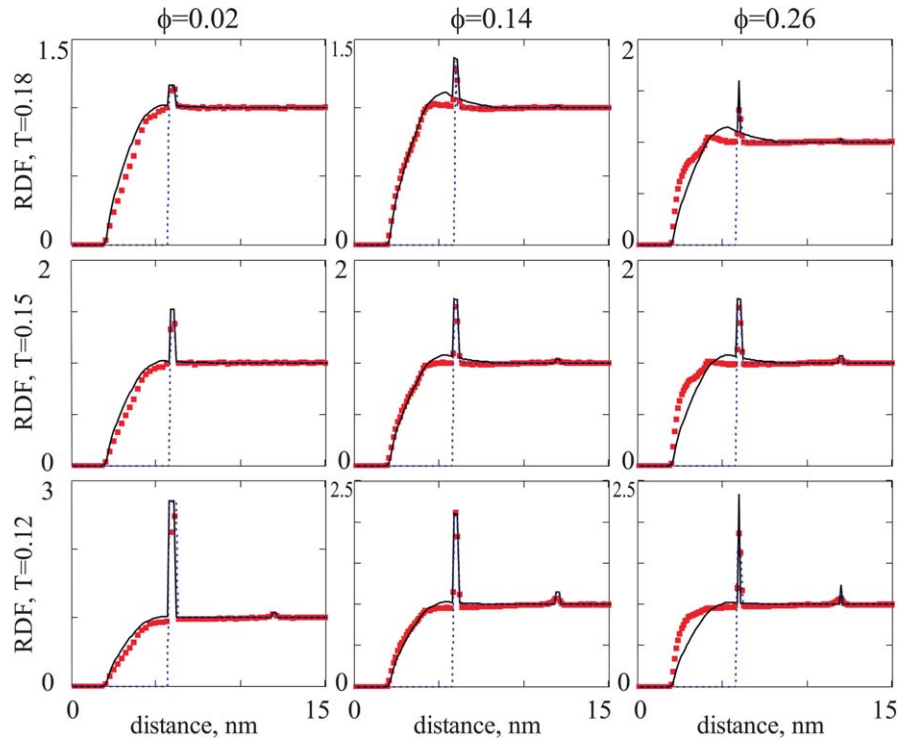

Fig. 5 Radial distribution functions for DNA self-assembling duplexes: $X_0 = 3$, symbols correspond to the simulation data; black solid lines describe the RDFs obtained theoretically using the approximated Gay–Berne potential together with the chain model; blue dashed lines are theoretical predictions, in which only chains are considered. The three rows correspond to three different temperatures: from the top to the bottom $T = 0.18$, $T = 0.15$ and $T = 0.12$. Each column corresponds to the fixed density: from the left to the right $\phi = 0.02$, $\phi = 0.14$, $\phi = 0.26$. The diameter of the SQ is fixed to 2 nm.

accurately. It happens because the density of single SQs is so high for this state point that the linear density expansion of the pair correlation function does not suffice to describe steric interactions between the monomers. However, at larger distances, theoretical RDFs agree really well with the simulation data for the complete range of parameters. We are now at the stage that the structure factors can be calculated and compared with simulation data.

4 Structure factors (SFs)

4.1 Calculation details

In order to benefit sufficiently from scattering experiments, one needs to find a reliable theoretical approach to interpret experimental results: on the one hand, the observed scattering images can be explained by multiple cluster topologies and on the other, the reconstruction of the radial distribution function from the scattering snapshot involves back-Fourier transformations of discrete data points, which has been shown to be a complex and unstable mathematical problem. The analytical relationship between the centre–centre SF $S(\boldsymbol{q})$ and the radial distribution function $g(\boldsymbol{r})$ can be written as

$$S(\boldsymbol{q}) = 1 + \rho\int \mathrm{d}\boldsymbol{r}[g(\boldsymbol{r}) - 1]\exp(-i\boldsymbol{q}\cdot\boldsymbol{r}), \qquad (14)$$

where $\rho = N/V = \phi/v_\mathrm{d}$ denotes the number density and V is the sample volume which is populated by N particles/molecules with volume v_d; $\boldsymbol{q}$ is the wave vector; the vector connecting the centres of two particles is denoted by $\boldsymbol{r}$. As emphasised above, the SF is a Fourier image of the RDF. In cases where the system under study is spherically symmetric, as in our case, when the DNA chains are forming an isotropic phase and the Gay–Berne potential is no more angular dependent due to the simplification introduced above, we can simplify the expression in eqn (14) by spherical averaging:

$$S(q) = 1 + \frac{4\pi\rho}{q}\int_0^\infty r[g(r) - 1]\sin(qr)\mathrm{d}r, \qquad (15)$$

where r and q are the moduli of $\boldsymbol{r}$ and $\boldsymbol{q}$ respectively. Here, the remaining integral corresponds to the averaging over all distances accessible in the system.

In Fig. 6–8 we plot the SFs with anisotropy analytically taken into account with solid lines, and without the anisotropy – in dashed lines.

4.2 Discussion

The first conclusion which follows from these figures (6–8) is that the developed theoretical approach (black solid lines) is able to describe the SFs for a broad range of ϕ, T and SQs anisotropy (X_0). The model does not agree with the simulation data for low q. This region of wave vectors corresponds to large distances in the system, thus our approach based on the assumption of the low density (the DFT discussed in Section 2 describes a gas of chains, in which steric inter-chain interactions are taken into account on the basis of mean-field-type approximation), results in the value of compressibility different

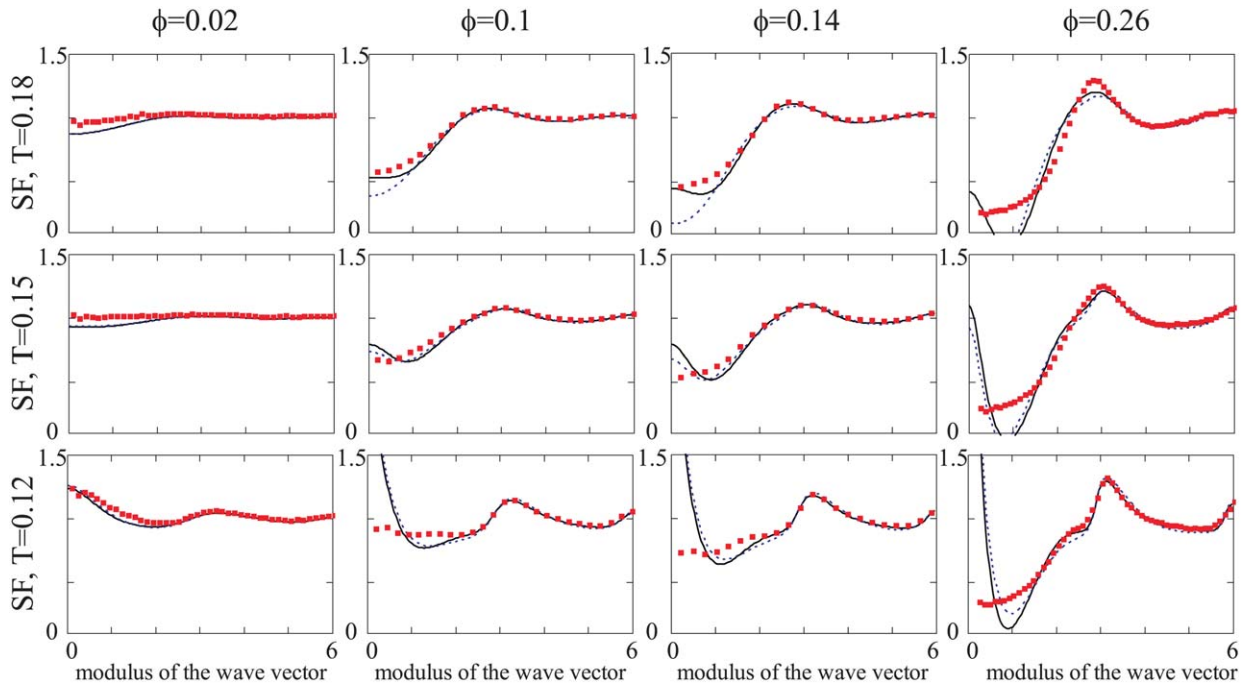

Fig. 6 SFs for the system of DNA self-assembling duplexes approximated by SQs: $X_0 = 1$, symbols correspond to the simulation data; black solid lines describe the SFs obtained theoretically using the approximated Gay–Berne potential together with the chain model for the RDF (Fig. 3, black line); blue dashed lines are theoretical predictions, in which only chains are considered (corresponding RDFs are plotted with blue dashed lines in Fig. 3). The three rows correspond to three different temperatures: from the top to the bottom $T = 0.18$, $T = 0.15$ and $T = 0.12$. Each column corresponds to the fixed density: from the left to the right $\phi = 0.02$, $\phi = 0.1$, $\phi = 0.14$, $\phi = 0.26$. The modulus of the wave vector is measured in nm^{-1}.

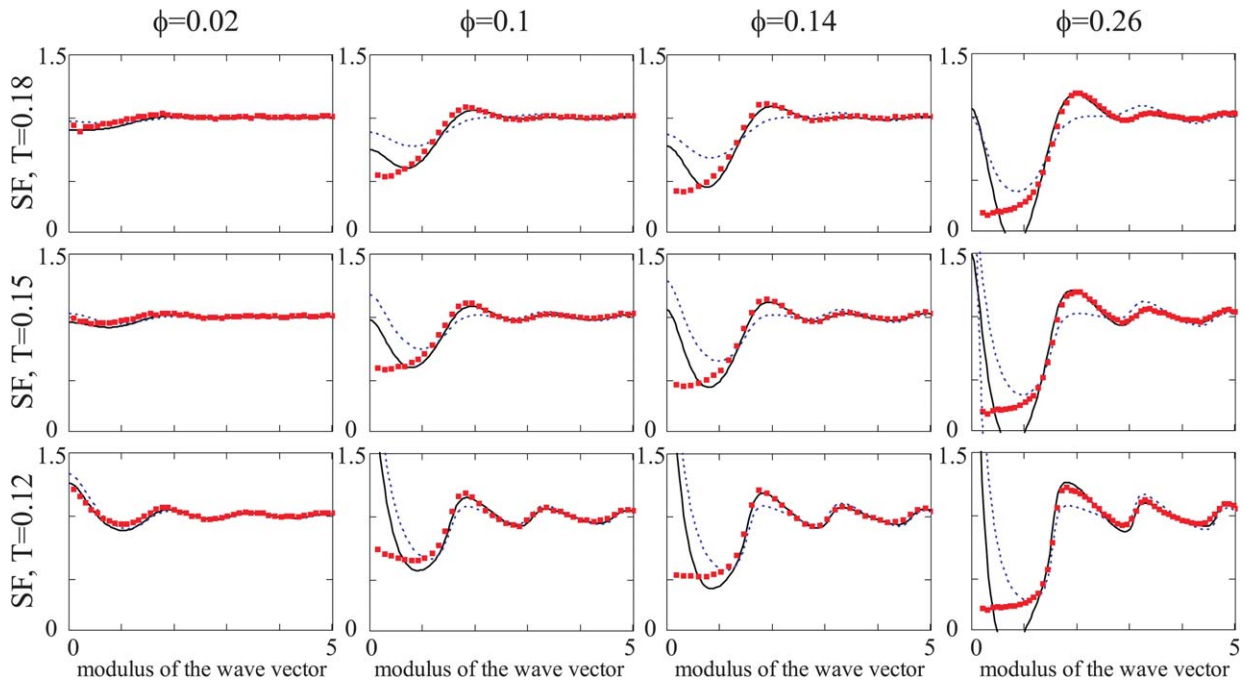

Fig. 7 SFs for the system of DNA self-assembling duplexes approximated by SQs: $X_0 = 2$, symbols correspond to the simulation data; black solid lines describe the SFs obtained theoretically using the approximated Gay–Berne potential together with the chain model for the RDF (Fig. 4, black line); blue dashed lines are theoretical predictions, in which only chains are considered (corresponding RDFs are plotted with blue dashed lines in Fig. 4). The three rows correspond to three different temperatures: from the top to the bottom $T = 0.18$, $T = 0.15$ and $T = 0.12$. Each column corresponds to the fixed density: from the left to the right $\phi = 0.02$, $\phi = 0.1$, $\phi = 0.14$, $\phi = 0.26$. The modulus of the wave vector is measured in nm^{-1}.

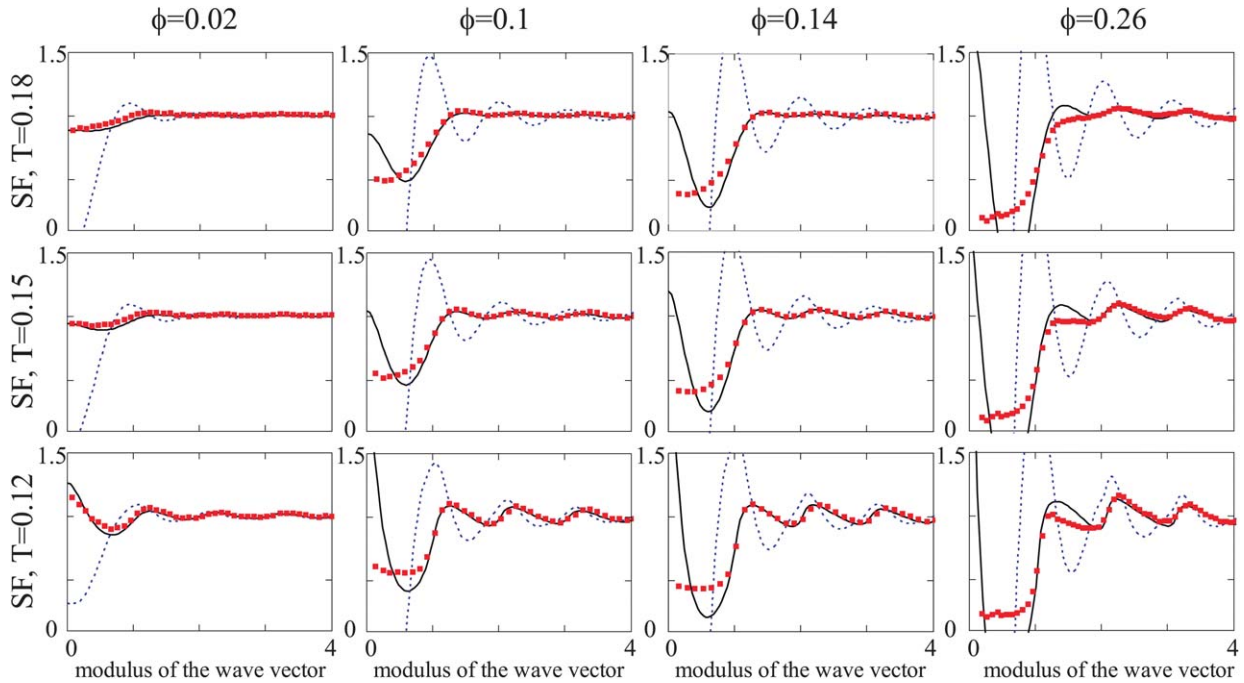

Fig. 8 SFs for the system of DNA self-assembling duplexes approximated by SQs: $X_0 = 3$, symbols correspond to the simulation data; black solid lines describe the SFs obtained theoretically using the approximated Gay–Berne potential together with the chain model for the RDF (Fig. 5, black line); blue dashed lines are theoretical predictions, in which only chains are considered (corresponding RDFs are plotted with blue dashed lines in Fig. 5). The three rows correspond to three different temperatures: from the top to the bottom $T = 0.18$, $T = 0.15$ and $T = 0.12$. Each column corresponds to the fixed density: from the left to the right $\phi = 0.02$, $\phi = 0.1$, $\phi = 0.14$, $\phi = 0.26$. The modulus of the wave vector is measured in nm^{-1}.

from the one obtained in MC simulations, where all possible steric interactions are accurately considered. However, as we will show later, in order to extract the average cluster size or the cluster size distribution for dilute systems, it is possible to investigate the SF first peak, its height, position and width. So from now on, we focus on these observables. One can see that for all three values of X_0, theory (black curves) and simulations (red symbols) agree well. In case of the highest density one can observe a slight deviation, which is related to the previously described discrepancies in the RDFs, due to the insufficiency of the first-order density expansion for the initial part of the pair correlation function for the anisotropic part. Similar to the plots of RDFs, the theoretical model, which does not consider the anisotropy of SQs (dashed blue lines), provides a reasonable agreement with the simulations only for very low values of ϕ and low values of X_0, meaning in the regions with no correlations or for quasi-isotropic particles.

In order to analyse the relationship between the width and height of the SF first peak and the chain size distribution we will use a theoretical model with anisotropy taken into account. We do not provide simulation results here for the sake of clarity, since they are in very good agreement with our theoretical predictions. In Fig. 9 we plot the position and the height of the first peak as a function of ϕ (the first and the second column, respectively). To evaluate the width of the first peak, plotted in the rightmost column of Fig. 9 as a function of ϕ, initially, the position of the first peak is fixed, then relative to this position the distance between the first on the left and the first on the right roots of the equation $S(q) = 1$ is calculated.

Firstly, it is clear that the anisotropy of SQs might qualitatively change the density dependence of the SF: if the anisotropy is high, the position of the first peak does not depend on ϕ; for moderate anisotropy the maximum shifts towards smaller q values; if the particles are quasi-isotropic, then the position of the first peak depends on the aggregation and can serve as a key for extracting the cluster size distribution. The height of the first peak for all systems investigated here grows with ϕ and is correlated with the amount of SQ pairs in the system. However in the case of highly anisotropic particles, instead of being unique the first peak of the structure factor might split into two (see Fig. 8) due to the presence of parallel side-by-side, "T"- and head-to-tail pairs, shown in Fig. 2.

Secondly, these figures (Fig. 9) might help to understand how active the chain formation is in the system. For that let us assume in an experiment, the measurements were performed at a given T and X_0, but at different densities. As a result of these series of measurements one has at hand the height, position and width of the first peak as functions of ϕ. Comparing experimental trends to the ones summarised in Fig. 10 one can conclude if the aggregation in the system is strong, average or weak. For example if the series of experiments were performed at $X_0 = 2$, then if the position of the SF stays constant, its height grows but the SF first peak becomes more and more narrow with growing density, according to the corresponding arrows, and the aggregation in the investigated system is weak. If, in contrast, the peak broadens, then the aggregation in the system is rather strong.

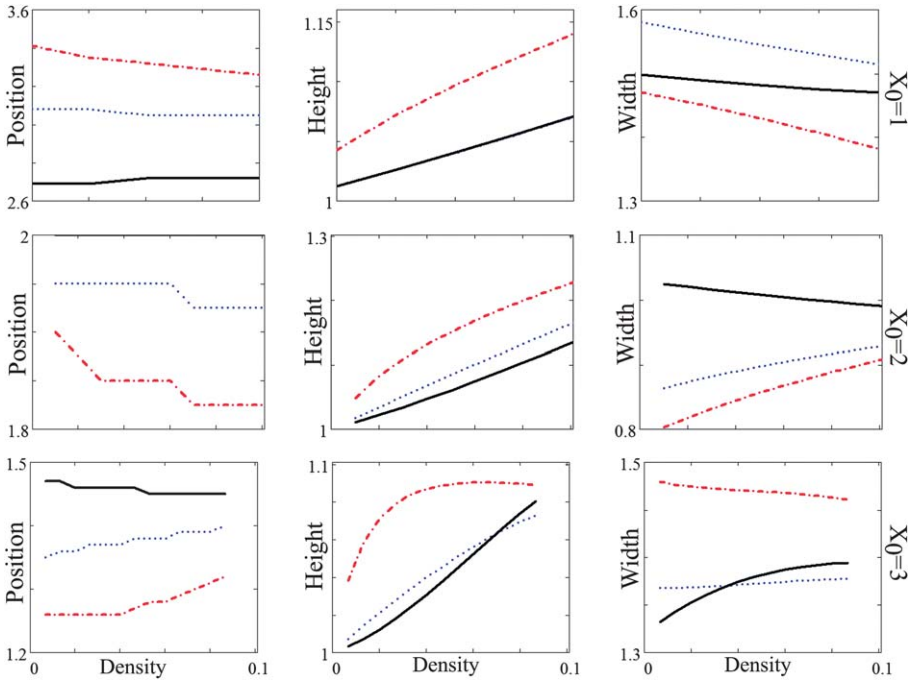

Fig. 9 Characteristics of the SF first peak. The first line corresponds to $X_0 = 1$; the second line is plotted for $X_0 = 2$ and the lowest row contains the plots for SQs with semiaxes ratio $X_0 = 3$. In the first column we present the position of the first maximum as a function of SQs' density; the second column contains the plots showing the height of the first peak as a function of SQs density; in the rightmost column one finds the width of the first peak as a function of SQs' density. In all nine plots, black lines describe theoretical predictions for $T = 0.18$; blue dashed lines are the theoretical predictions for $T = 0.15$; and curves corresponding to $T = 0.12$ are plotted in red dashed-dotted lines. Simulation results being in very good agreement with our theory are not provided here for the sake of clarity.

Thirdly, what follows from the figure is the following. If one fixes ϕ and X_0, but measures the SFs for building blocks with various functionalities (or various interaction energies), then the following rule holds true: the longest chains are there in the system with $X_0 = 1$, if the first structure factor maximum has the rightmost position, the highest peak, and this peak is the narrowest; if $X_0 = 2$, the leftmost peak position, the highest and the narrowest peak would correspond to the longest chains; for the highest value $X_0 = 3$ the system with the longest chains will be described by the leftmost peak position with both maximal height and width. This rule is graphically summarised in Fig. 11.

In a more detailed way, one can say that the position of the first peak corresponds to the characteristic radius of the first coordination sphere in the system. Its shift to the right (towards larger wave vectors) means the decrease of the latter radius. So the chain formation in the system with $X_0 = 1$, for example, results in a larger characteristic distance for $T = 0.12$, and in a smaller characteristic distance for $T = 0.18$. It can be easily understood from the chain length dependence provided in the

X_0	Position	Height	Width	Chain formation
$X_0 = 1$	←	↑	↓	*Active*
	↔	↑	↓	*Moderate*
	→	↑	↓	*Weak*
$X_0 = 2$	↔	↑	↑	*Active*
	↔	↑	↓ ↔	*Moderate*
	↔	↑	↓	*Weak*
$X_0 = 3$	↔	↑	↓	*Active*
	→ ↔	↑	↔	*Moderate*
	↔ split	↑	↑	*Weak*

Fig. 10 The table shows the density effect on the characteristic behaviour of the position, height and width of the structure factor first peak for $X_0 = 1, 2, 3$. The arrow pointing up corresponds to the increase of the observable with ϕ, the arrow pointing down characterises the observable as a decreasing function of ϕ, the arrow pointing to the right means the shift of the peak position towards larger wave vectors and, finally, the double-sided right–left arrow symbolises negligible changes of the observable with ϕ. In the last row we provide the estimate for the self-assembly intensity.

X_0	Position	Height	Width
$X_0 = 1$	→	↑	↓
$X_0 = 2$	←	↑	↓
$X_0 = 3$	←	↑	↑

Fig. 11 The table shows the effect of aggregation at fixed ϕ on the position, height and width of the SFs first peak for $X_0 = 1, 2, 3$. The arrow pointing to the right (to the left) means that the more the peak is shifted to the right (to the left), the stronger is the aggregation; the arrow pointing up (down) means that the largest (smallest) value of the corresponding observable would correspond to the longest chains.

left column of Fig. 1, where the increase of the density and decrease of the temperature drives to a significant chain length growth. For $X_0 = 1$, when the chains are formed, the interparticle distance becomes on average smaller than for the system of isotropically distributed SQs. When the anisotropy of particles comes into play, the interpretation becomes more complex. Even though the chain growth would still correspond to the decrease of the distance between particles within the chains, there are smaller characteristic distances related to "T"- and side-by-side configurations (see Fig. 2), whose contribution will become more and more pronounced with the growing value of single particles (see Fig. 1, second column). This competition results in a rather weak dependence of the first peak on the density of SQs for the systems with high anisotropy. Note that this observation would have been impossible, had the chain-only model been used without anisotropy of the SQs taken into account.

For any anisotropy, the height of the SF first peak grows with density (second column, Fig. 9). This growth is more pronounced when the anisotropy is $X_0 = 2$ (middle row) and can be on the order of 12% for the lowest temperature and high density (red dashed dotted line), whereas for $X_0 = 1$ and $X_0 = 3$ this increase does not exceed 5%. Independent of the SQs anisotropy, the higher the temperature is, the less the height grows (compare red dashed dotted lines, corresponding to $T = 0.12$ to black solid lines describing the height of the first peak for $T = 0.18$, Fig. 9, middle row). The height of the first peak for a simple isotropic system is related to the number of pairs at a corresponding aforementioned characteristic distance. In other words, the higher the chain length is, or the higher the simple probability of finding two particles close to each other is, the higher is the peak of the structure factor. The anisotropy here leads to the following effect: for $X_0 = 1$ the highest peak is observed, as long as independently from the orientation, be it "T"-, side-by-side or head-to-tail configuration, all contribute to the first coordination sphere, whereas for $X_0 = 3$ the height of the structure factor first peak is the smallest, due to the fact that coordination spheres have a high polydispersity of radii. In this way the high particle anisotropy can even lead to the split of the SF first peak into two, see the third column in Fig. 8.

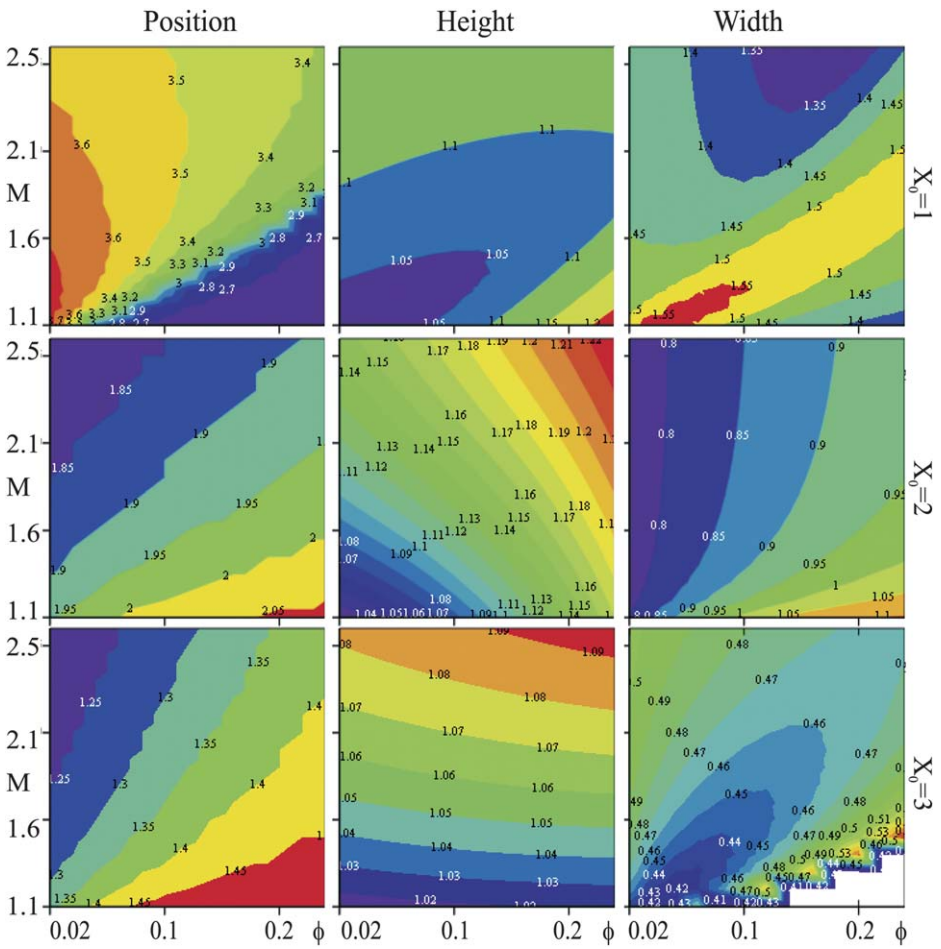

Fig. 12 M–ϕ colour diagrams. From the top to the bottom the SQs anisotropy grows: $X_0 = 1$ – first row; $X_0 = 2$ – second row; $X_0 = 3$ – third row. In all diagrams we plot ϕ along the abscissa axes and M along the ordinate axes. In the first row the numbers and the colours stay for the SF first peak position; the second row contains the information about the height of the SF first peak; in the third row the corresponding SF first peak width is presented. The white corner in the bottom-right figure corresponds to the area of parameters for which the SF first peak splits into two.

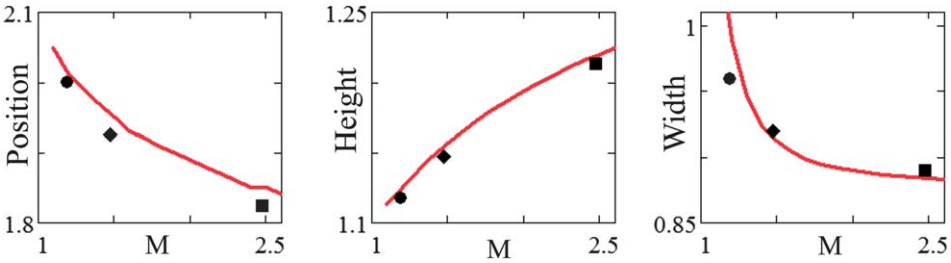

Fig. 13 Characteristics of the SF first peak as functions of the average chain length M for $\phi = 0.18$ and $X_0 = 2$: the left most plot is the position of the peak; the middle one is the height of the peak; the rightmost plot shows the width of the peak. Curves correspond to the theory developed in this study; points are taken from the simulations at the same state point and serve to verify the accuracy of the method. The agreement proves to be within 5%.

The main physical reason for the SF first peak width to change is the evolution of the range of probable and accessible distances in the system. Thus, for example, for isotropic monodisperse particles, intensive anisotropic attraction, leading to the chain formation, results in the fact that two particles might be found at a close contact with growing probability. This distance becomes dominant in the system and, thus, defines the characteristic wave vector corresponding to the fist peak.[32] For polydisperse particles forming chains, the situation is different:[33] the presence of several close-contact distances corresponding to the contacts between particles of different sizes leads to the appearance of a range of wave vectors, thus making the structure factor first peak broader. If one looks at the rightmost column in Fig. 9, the width of the first peak might both increase and decrease as a function of SQs density. In the result, the peak width decreases for $X_0 = 1$ and is almost constant for $X_0 = 3$, albeit for the highest temperature, for which the width slightly increases (due to increasing number of highly anisotropic monomers).

To summarise the recipe put forward in this manuscript, we conclude this section with two figures. Fig. 12 contains nine M–ϕ diagrams. In each diagram the abscissa shows the density, and the ordinate corresponds to the average chain length. The colour and the numbers in the diagram body in the first column correspond to the SF first peak position. The second column contains the information about the height of the peak; the third column diagrams show the width of the SF first peak. How can we trace back the cluster size distribution by analysing these characteristics of the structure factor first peak? We suggest making the following steps.

• Measuring the three main characteristics of the SF first peak, knowing the density and the building block anisotropy.

• Locating the experimental points on the diagrams from Fig. 12, and collecting the information about the average chain length.

• Using the framework given by eqn (2)–(8), calculating the cluster size distribution the free energy of interaction (*i.e.* the bonding free energy) between the building blocks. The latter could be calculated by finding the argument of the exponent in the expression for M from (8).

To check the accuracy of the method, in Fig. 13 we plot the position (the left most picture), height (middle) and width (rightmost) as functions of the average cluster size M for fixed $\phi = 0.18$ and $X_0 = 2$. Curves are obtained using the theoretical approach developed here; as benchmarks we also use the simulation data for the corresponding state point. One can see that the accuracy of the theoretical method lies within 5%.

5 Conclusion

In this work we develop a general theoretical approach to describe the structure factor of prolate cylindrically or elliptically symmetric particles which self-assemble into linear chains. Our approach is based on the direct analytical calculation of the centre-to-centre radial distribution function. To construct such a function we consider two main contributions: the one arising from sterically interacting anisotropic monomers and the one arising from chain formation.

Anisotropic monomers are modelled *via* a simple parameterisation for the Gay–Berne potential. Such a parameterisation allows us to calculate analytically a density expansion of the pair correlation. It turns out that already the first-order expansion provides an accurate modelling for a wide range of temperature, anisotropy and density. Chains contribute to the radial distribution function according to the model developed in ref. 32–34, where the relevant intra-chain centre-to-centre distances were calculated directly from the chain length distribution. The latter is determined here by free energy minimisation of a suitable free-energy functional.[22] With the centre-to-centre radial distribution function at hand, we calculate the structure factor of the isotropic system by a simple Fourier transform.

We test our approach on a simple coarse-grained model of short DNA duplex self-assembly.[22] We thoroughly investigate the behaviour of the model for various degrees of anisotropy, interaction strengths and densities and find a very good agreement with simulation data for a broad range of wave vectors, excluding the small wave vectors region. Despite such a limitation, the theory is able to provide a quantitative description of the degree of aggregation in the system.

Our theory assumes a specific distribution of chain length. If structures different from chains are present or if the chain distribution is not exponential, the interpretation of the scattering pattern requires additional investigation. But, under the assumption that the chain length distribution is exponential, two important pieces of information can be extracted by comparing the scattering pattern with the theoretical

predictions, namely (i) the average chain length M (*i.e.* the degree of association) and (ii) the stacking free energy G_{ST} associated with the formation of a bond. The average chain length M can be calculated by analysing the first peak of the structure factor (*i.e.* by measuring its width, height or position). The value of G_{ST} can be calculated from M using the relationship provided by the theory (see eqn (8)).

The final message of our work is that the anisotropy of the building blocks strongly affects the scattering behaviour of self-assembling systems. However, the specific signature of the latter anisotropy allows us to quantify the microscopic picture of the system, by interpreting the characteristics of the structure factor first peak. Currently, we are working on the extension of the present framework in order to describe the scattering patterns of partially ordered phases formed by polydisperse chains.

Appendix 1

In this part we describe how to calculate the pair correlation function for a dilute system of anisotropic prolate SQs. One of the widely used steric short range potentials which depends not only on the distance between particle centres, but also on their orientation is a Gay–Berne potential.[45,46,48] The original version of this potential has the functional form of the Lennard Jones potential.[49] Here we need only the repulsive part of the potential, and for that one could write a Weeks–Chandler–Andersen[47] modification of the Gay–Berne potential:

$$U_{GB}(\boldsymbol{u}_i,\boldsymbol{u}_j,\boldsymbol{r}_{ij}) = \begin{cases} 4\varepsilon(\,)[A^{12}(\,)-A^{6}(\,)]+\varepsilon(\,), & r_{ij}\le r_c \\ 0, & r_{ij}>r_c, \end{cases} \quad (16)$$

where

$$A(\,)\equiv A\left(\boldsymbol{u}_i,\boldsymbol{u}_j,\hat{\boldsymbol{r}}_{ij}\right)=\sigma_0\Big/\left(r_{ij}-\sigma\left(\boldsymbol{u}_i,\boldsymbol{u}_j,\hat{\boldsymbol{r}}_{ij}\right)+\sigma_0\right),\sigma\left(\boldsymbol{u}_i,\boldsymbol{u}_j,\hat{\boldsymbol{r}}_{ij}\right)$$
$$=\sigma_0\left[1-\frac{\chi(X_0)}{2}\right.$$
$$\left.\times\left\{\frac{\left(\hat{\boldsymbol{r}}_{ij}\cdot\boldsymbol{u}_i+\hat{\boldsymbol{r}}_{ij}\cdot\boldsymbol{u}_j\right)^2}{1+\chi(X_0)\ \boldsymbol{u}_i\cdot\boldsymbol{u}_j}+\frac{\left(\hat{\boldsymbol{r}}_{ij}\cdot\boldsymbol{u}_i-\hat{\boldsymbol{r}}_{ij}\cdot\boldsymbol{u}_j\right)^2}{1-\chi(X_0)\ \boldsymbol{u}_i\cdot\boldsymbol{u}_j}\right\}\right]^{-\frac{1}{2}};$$
$$\varepsilon(\,)\equiv\varepsilon\left(\boldsymbol{u}_i,\boldsymbol{u}_j\right)=\varepsilon_0\left[1-\chi^2(X_0)\left(\boldsymbol{u}_i\cdot\boldsymbol{u}_j\right)^2\right]^{-\frac{1}{2}}.$$

here, $\hat{\boldsymbol{r}}=\boldsymbol{r}_{ij}/r_{ij}=(\boldsymbol{r}_i-\boldsymbol{r}_j)/r_{ij}$ is the unit vector, coaligned with the vector connecting particle centres, $\boldsymbol{u}_{i(j)}$ is the unit vector along the principal axis of the particle, $\sigma_0=\sqrt{2}D$, where D is the diameter of the duplex, ε_0 denotes the energy parameter, and $\chi(X_0)=[X_0^2-1]/[X_0^2+1]$. The critical value r_c equals to $(2^{1/6}-1)\sigma_0+\sigma(\boldsymbol{u}_i,\boldsymbol{u}_j,\hat{\boldsymbol{r}}_{ij})$. It is important to underline that this potential is not central and depends on mutual orientation of particles through, for example, the effective interparticle distance $\sigma(\,)$, which depends on the SQs orientations.

In order to calculate $\beta_2(r,T,X_0)$, we fix the coordinate system in such a way that the centre of mass of the first SQ is in the origin and the z axis is coaligned with $\boldsymbol{u}_1$. In this coordinate frame, $\beta_2(r,T,X_0)$ has the following form:

$$\beta_2(r,T,X_0) = \frac{1}{16\pi^2}\int_0^{2\pi}\int_0^{2\pi}\int_0^{\pi}\int_0^{\pi}\sin\theta\sin\zeta \times\exp\left(-\frac{U_{GB}(\boldsymbol{u}_1,\boldsymbol{u}_2,\boldsymbol{r})}{k_BT}\right)d\zeta d\theta d\varphi d\xi.$$

Using the functional for the potential, one of the polar angles can be integrated out resulting in a factor 2π in front of the integral:

$$\beta_2(r,T,X_0) = \frac{1}{8\pi}\int_0^{2\pi}\int_0^{\pi}\int_0^{\pi}\sin\theta\sin\zeta\times\exp\left(-\frac{U_{GB}(\boldsymbol{u}_1,\boldsymbol{u}_2,\boldsymbol{r})}{k_BT}\right). \quad (17)$$

The triple integral, being not easy to calculate, could have been of no impediment by itself, but in the present form higher order coefficients will be expressed *via* nine and more multifold integrals, thus making the analytical construction of the RDF far too costly. In order to solve this problem, we propose to approximate the anisotropic Gay–Berne potential by an isotropic one, using the parameterisation of the effective interparticle distance $\sigma(\,)$. As the fitting criteria we chose in this work the dependence of β_2 (see eqn (17)) on the distance. In other words, the approximated potential should exhibit the same zero-order pair correlation function as the one with the nonparameterised distance. This choice, besides having a thermodynamics ground, is also very convenient as it involves only one time calculation of the triple integral. The parameterisation found in this way leads to the following modified isotropic Gay–Berne potential:

$$U_{GB}^{apr}(t,r_{ij}) = \begin{cases} 4\varepsilon(\,)[A^{12}(\,)-A^{6}(\,)]+\varepsilon(\,), & r_{ij}\le r_c \\ 0, & r_{ij}>r_c, \end{cases} \quad (18)$$

$$A(\,)\equiv A(t,r_{ij})=\sigma_0/(r_{ij}-\sigma(t)+\sigma_0)$$
$$\sigma(t)=\sigma_0[1-t\chi(X_0)/(1-\chi^2(X_0))]^{-1/2};$$
$$\varepsilon(\,)=\varepsilon_0[1-\chi^2(X_0)]^{-1/2},\ r_c=(2^{1/6}-1)\sigma_0+\sigma(t).$$

Unfortunately, finding a unique parametrisation t for any value of the anisotropy parameter X_0 turns out to be very complicated. In this study, we unify the range of the parameterisation s and fix t to have a polynomial functional form: $t\equiv t(s,X_0)$. Under this assumption, the expression for β_2 assumes a short simple form:

$$\beta_2(r,T,X_0) = \frac{1}{0.86}\int_{-0.01}^{0.85}\exp\left(-\frac{U_{GB}^{apr}(t(s,X_0),r)}{k_BT}\right)ds, \quad (19)$$

where

$$t(s,X_0)=\sum_{i=1}^{7}(-1)^{i+1}a_i(X_0)s^i,$$

for $i\neq 2$ $a_i(X_0)>0$ and $a_2(X_0)\equiv 0$.

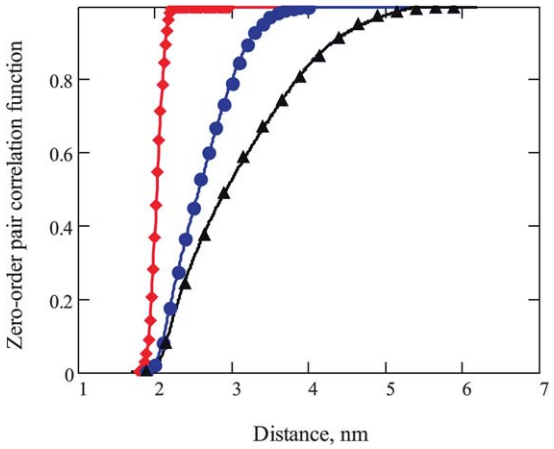

Fig. 14 The zero-order pair correlation function for anisotropic monomers as a function of the centre–centre distance. Approximations from eqn (17) are shown with solid lines; the exact value from eqn (19) is plotted in symbols. Different colours represent different semiaxes ratios: red denotes $X_0 = 1$, blue $X_0 = 2$ and black ones are for $X_0 = 3$.

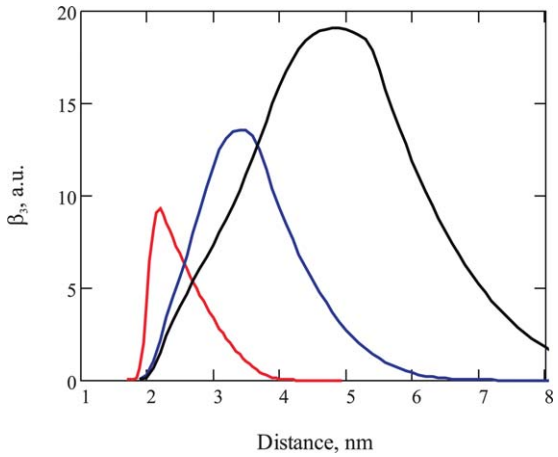

Fig. 15 Three particle contributions to the pair correlation function for anisotropic monomers as a function of the centre–centre distance. Approximations from eqn (20) are shown with solid lines. Different colours represent different semiaxes ratios: red denotes $X_0 = 1$, blue $X_0 = 2$ and black ones are for $X_0 = 3$.

We find the values of $a_i(X_0)$, as it was mentioned above, by fitting the values of the approximated β_2 to the ones for the non-modified Gay–Berne potential. The results are presented in Fig. 14, where β_2 is plotted as a function of the centre–centre distance measured in D. The values corresponding to the exact zero-order correlation function eqn (17) are plotted with solid lines and results obtained using eqn (19) are presented in symbols. One can see very good agreement between solid lines and symbols. The deviation here does not exceed 2%. The behaviour of β_2 is determined by the anisotropy ratio, and its shape gives rise to the shoulder we observed in the simulation data (see Fig. 3–5). All curves start from the distance equal to two particle radii, which corresponds to the side-by-side orientation, and the centre–centre distance in this case is identical for all three semiaxes ratios. Now, we can use our modified Gay–Berne potential eqn (18) for calculating β_3 and higher order coefficients. In this study, we exploit the fact that the density of monomers is rather low, so we expand the RDF only up to the linear term. For that one needs to calculate β_3:

$$\beta_3(r, T, X_0) = 2\pi\beta_2(r, T, X_0) \int_{-0.01}^{0.85}\int_{-0.01}^{0.85}\int_{0}^{r_c} r_{13}^2 \left[f(s_{13}, r) \times \left(\int_{-1}^{1} \left(f\left(s_{23}, \sqrt{(r_{13}^2 + r^2 - 2rr_{13}q)} - 1 \right) \right) \mathrm{d}q - 2 \right) \right] \mathrm{d}r_{13}\mathrm{d}s_{13}\mathrm{d}s_{23}. \quad (20)$$

To provide a picture of how the three-particle steric contributions look like, we plot β_3 in the following plot (Fig. 15).

In this figure one sees that the maximum of all these curves is between T and head-to-tail configurations, which leads to a pronounced first minimum of the RDF observed in simulations (see Fig. 3–5). We used the aforementioned approach to calculate the total RDF, see eqn (13).

Appendix 2

In this section we provide the explanation of how to obtain RDF in the form described in eqn (12). The full radial distribution function of a system of N particles is by definition:

$$g(\boldsymbol{r}_1, \boldsymbol{r}_2) = \frac{N!}{(N-2)!}\frac{1}{Z_N}\int \mathrm{e}^{-\beta U_N}\mathrm{d}\boldsymbol{r}^{(N-2)}, \quad (21)$$

where Z_N is the configurational integral (see ref. 43 pp. 28–30) and the sum of all interparticle interactions is

$$U_N = \sum_{ij} u(\boldsymbol{r}_i, \boldsymbol{r}_j). \quad (22)$$

Here, $u(\boldsymbol{r}_i, \boldsymbol{r}_j)$ stays for the total interaction of particles i and j. We consider now a mixture of monomers and chains of length $l > 1$ (which will be called simply "chains" in the following); the radial distribution function (RDF) of this system can be written as follows:

$$g(\boldsymbol{r}_1, \boldsymbol{r}_2, T, X_0, \phi, \delta) = x_m x_m g_{mm}(\boldsymbol{r}_1, \boldsymbol{r}_2, T, X_0, \phi) + 2x_m x_c g_{mc}(\boldsymbol{r}_1, \boldsymbol{r}_2, T, X_0, \phi) + x_c x_c g_{cc}(\boldsymbol{r}_1, \boldsymbol{r}_2, T, X_0, \phi, \delta), \quad (23)$$

where m stands for monomers, c for chains and if $\alpha, \beta \in \{m, c\}$ $g_{\alpha\beta}(\,)$ are partial RDFs and $x_m = \nu(1)/\rho$, $x_c = (\rho - \nu(1))/\rho$.

The expression for $g_{mm}(\,)$ can be expanded in powers of the density ρ as follows,[43] up to the first order:

$$g_{mm}(r, T, X_0, \phi) \approx \beta_2(r, T, X_0) + \nu(1)\beta_3(r, T, X_0), \quad (24)$$

where $\beta_2(r, T, X_0)$ and $\beta_3(r, T, X_0)$ are calculated according to a Gay–Berne interaction potential between monomers (see

Appendix 1). In the expression eqn (24), the RDF depends on the interparticle distance r only. The expression for g_{mc} has the form:

$$g_{mc}(\boldsymbol{r}_m, \boldsymbol{r}_c, T, X_0, \phi) = 2\frac{N_c!}{(N_c-1)!}\frac{N_m!}{(N_m-1)!}\frac{1}{2x_m x_c \rho^2 Z_N} \times \int e^{-\beta U_N} d\boldsymbol{r}^{(N-2)} = \frac{N^2}{\rho^2 Z_N}\int e^{-\beta U_N} d\boldsymbol{r}^{(N-2)}. \quad (25)$$

Now if we introduce the Mayer function, *i.e.*

$$f(\boldsymbol{r}_i,\boldsymbol{r}_j) = e^{-\beta u(\boldsymbol{r}_i,\boldsymbol{r}_j)} - 1, \quad (26)$$

and perform a classical cluster expansion, we obtain:

$$g_{mc}(\boldsymbol{r}_m, \boldsymbol{r}_c, T, X_0, \phi) = e^{-\beta u(\boldsymbol{r}_m,\boldsymbol{r}_c)} V^2 V^{N-2} \times \left[1 + \frac{N_c}{V}\int f(\boldsymbol{r}_m, \boldsymbol{r}_{c'}) f(\boldsymbol{r}_{c'}, \boldsymbol{r}_c) d\boldsymbol{r}_{c'} + \frac{N_m}{V}\int f(\boldsymbol{r}_m, \boldsymbol{r}_{m'}) f(\boldsymbol{r}_{m'}, \boldsymbol{r}_c) d\boldsymbol{r}_{m'} + o(\rho)\right] = e^{-\beta u(\boldsymbol{r}_m,\boldsymbol{r}_c)}\left[1 + \frac{N_c}{V}\int f(\boldsymbol{r}_m, \boldsymbol{r}_{c'}) f(\boldsymbol{r}_{c'}, \boldsymbol{r}_c) d\boldsymbol{r}_{c'} + \frac{N_m}{V}\int f(\boldsymbol{r}_m, \boldsymbol{r}_{m'}) f(\boldsymbol{r}_{m'}, \boldsymbol{r}_c) d\boldsymbol{r}_{m'}\right], \quad (27)$$

where Z_N is the configurational integral, N_m (N_c) in front of the integral accounts for the possibility to choose any monomer $\boldsymbol{r}_{m'}$ (chain $\boldsymbol{r}_{c'}$) to integrate over and we used the fact that:

$$Z_N = V^N[1 + V\rho^2\beta_2(r, T, X_0) + o(\rho^2)]. \quad (28)$$

If we neglect in eqn (27) $o(\rho)$ terms and the term in N_c, which accounts for chain–chain interactions, we have:

$$g_{mc}(\boldsymbol{r}_m, \boldsymbol{r}_c, T, X_0, \phi) = e^{-\beta u(\boldsymbol{r}_m,\boldsymbol{r}_c)} \times [1 + \nu(1)\int f(\boldsymbol{r}_m, \boldsymbol{r}_{m'}) f(\boldsymbol{r}_{m'}, \boldsymbol{r}_c) d\boldsymbol{r}_{m'}] \approx \beta_2(r, T, X_0) + \nu(1)\beta_3(r, T, X_0), \quad (29)$$

where after simplification, the dependence on the two position vectors can be replaced by the dependence on the interparticle distance. $g_{cc}(r)$ is approximated as follows:

$$g_{cc}(r, T, X_0, \phi, \delta) = \beta_2(r, T, X_0) + \sum_s p^s / A_s(X_0, \phi, \delta, r), \quad (30)$$

where the Gay–Berne contribution to the RDF is accounted for only at the level of steric hindrance through the β_2 term. The 3-body first order contribution proportional to β_3 is dropped because for monomers aggregated into linear chains this type of interaction is not crucial.

Acknowledgements

The research was carried out with the financial support of FP7-IDEAS-ERC grant PATCHYCOLLOIDS. One of the authors (E.P.) is grateful to the Ural Federal University development program with the financial support of young scientists. S.K. and E.P. were supported by RFBR grant mol-a 1202-31-374, RFBR grant mol-a-ved 12-02-33106 and Ministry of Science and Education of RF 2.609.2011.

References

1 M. Nakata, G. Zanchetta, B. D. Chapman, C. D. Jones, J. O. Cross, R. Pindak, T. Bellini and N. A. Clark, *Science*, 2007, **318**, 1276.
2 G. Zanchetta, F. Giavazzi, M. Nakata, M. Buscaglia, R. Cerbino, N. A. Clark and T. Bellini, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 17497–17502.
3 J. Lydon, *Curr. Opin. Colloid Interface Sci.*, 1998, **3**, 458–466.
4 J. Lydon, *J. Mater. Chem.*, 2010, **20**, 10071–10099.
5 F. Chami and M. R. Wilson, *J. Am. Chem. Soc.*, 2010, **132**, 7794–7802.
6 P. van der Schoot and M. Cates, *Europhys. Lett.*, 1994, **25**, 515–520.
7 A. Khan, *Curr. Opin. Colloid Interface Sci.*, 1996, **1**, 614.
8 D. M. Kuntz and L. M. Walker, *Soft Matter*, 2008, **4**, 286–293.
9 C. F. Lee, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2009, **80**, 031902.
10 J.-M. Jung and R. Mezzenga, *Langmuir*, 2010, **26**, 504–514.
11 R. Mezzenga, J.-M. Jung and J. Adamcik, *Langmuir*, 2010, **26**, 10401–10405.
12 P. Mariani, F. Spinozzi, F. Federiconi, H. Amenitsch, L. Spindler and I. Drevensek-Olenik, *J. Phys. Chem. B*, 2009, **113**, 7934–7944.
13 S. Lena, G. Brancolini, G. Gottarelli, P. Mariani, S. Masiero, A. Venturini, V. Palermo, O. Pandoli, S. Pieraccini, P. Samorì and G. Spada, *Chem.–Eur. J.*, 2007, **13**, 3757–3764.
14 I. Budin and J. W. Szostak, *Annu. Rev. Biophys.*, 2010, **39**, 245–263.
15 T. de Lange, *Science*, 1998, **279**, 334–335.
16 W. C. Hahn, C. M. Counter, A. S. Lundberg, R. L. Beijersbergen, M. W. Brooks and R. A. Weinberg, *Nature*, 1999, **400**, 464–468.
17 J. T. Davis, *Angew. Chem., Int. Ed.*, 2004, **43**, 668–698.
18 E. S. Baker, J. T. Lee, J. L. Sessler and M. T. Bowers, *J. Am. Chem. Soc.*, 2006, **128**, 2641–2648.
19 Y. Olivier, L. Muccioli, V. Lemaur, Y. H. Geerts, C. Zannoni and J. Cornil, *J. Phys. Chem. B*, 2009, **113**, 14102–14111.
20 S. V. Shiyanovskii, T. Schneider, I. I. Smalyukh, T. Ishikawa, G. D. Niehaus, K. J. Doane, C. J. Woolverton and O. D. Lavrentovich, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2005, **71**, 020702.
21 A. Aguzzi, *Nature*, 2009, **459**, 924–925.
22 C. De Michele, T. Bellini and F. Sciortino, *Macromolecules*, 2012, **45**, 1090–1106.
23 C. De Michele, L. Rovigatti, T. Bellini and F. Sciortino, *Soft Matter*, 2012, **8**, 8388–8398.
24 C. De Michele, *J. Comput. Phys.*, 2010, **229**, 3276–3294.
25 B. Kezic and A. Perera, *J. Chem. Phys.*, 2011, **135**, 234104.
26 Y. Duda, C. J. Segura, E. Vakarin, M. F. Holovko and W. G. Chapman, *J. Chem. Phys.*, 1998, **108**, 9168–9176.

27 A. Yethiraj and C. K. Hall, *J. Chem. Phys.*, 1990, **93**, 5315–5321.
28 D. Pearlman, D. Case, J. Caldwell, W. Ross, T. Cheatham III, S. DeBolt, D. Ferguson, G. Seibel and P. Kollman, *Comput. Phys. Commun.*, 1995, **91**, 1–41.
29 A. D. Mackerell and M. K. J. Wiorkiewicz Kuczera, *J. Am. Chem. Soc.*, 1995, **117**, 11946–11975.
30 A. Morriss-Andrews, J. Rottler and S. S. Plotkin, *J. Chem. Phys.*, 2010, **132**, 035105.
31 T. E. Ouldridge, I. G. Johnston, A. A. Louis and J. P. K. Doye, *J. Chem. Phys.*, 2009, **130**, 065101.
32 A. Ivanov, S. Kantorovich and E. Pyanzina, *Magnetohydrodynamics*, 2007, **43**, 33–38.
33 E. Pyanzina, S. Kantorovich, J. J. Cerda, A. Ivanov and C. Holm, *Mol. Phys.*, 2009, **107**, 571–590.
34 E. Pyanzina, S. Kantorovich, J. Cerda and C. Holm, *J. Magn. Magn. Mater.*, 2011, **323**, 1263–1268.
35 C. T. Lee, K. A. Smith and T. A. Hatton, *Langmuir*, 2009, **25**, 13784–13794.
36 J. H. Ortony, T. Chatterjee, L. E. Garner, A. Chworos, A. Mikhailovsky, E. J. Kramer and G. C. Bazan, *J. Am. Chem. Soc.*, 2011, **133**, 8380–8387.
37 C. De Michele, S. Gabrielli, P. Tartaglia and F. Sciortino, *J. Phys. Chem. B*, 2006, **110**, 8064–8079.
38 C. De Michele, P. Tartaglia and F. Sciortino, *J. Chem. Phys.*, 2006, **125**, 204710.
39 S. Corezzi, C. De Michele, E. Zaccarelli, D. Fioretto and F. Sciortino, *Soft Matter*, 2008, **4**, 1173–1177.
40 S. Corezzi, C. De Michele, E. Zaccarelli, P. Tartaglia and F. Sciortino, *J. Phys. Chem. B*, 2009, **113**, 1233–1236.
41 T. Kouriabova, M. Betterton and M. Glaser, *J. Mater. Chem.*, 2010, **20**, 10366–10383.
42 X. Lü and J. Kindt, *J. Chem. Phys.*, 2006, **125**, 054909.
43 J.-P. Hansen and I. McDonald, *Theory of Simple Liquids*, Academic Press, London, 3rd edn, 1986.
44 R. Balescu, *Equilibrium and Nonequilibrium Statistical Mechanics*, Krieger, Malabar, FL, 1st edn, 1991.
45 B. J. Berne and P. Pechukas, *J. Chem. Phys.*, 1972, **56**, 4213–4216.
46 D. J. Cleaver, C. M. Care, M. P. Allen and M. P. Neal, *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.*, 1996, **54**, 559–567.
47 J. D. Weeks, D. Chandler and H. C. Andersen, *J. Chem. Phys.*, 1971, **54**, 5237.
48 J. G. Gay and B. J. Berne, *J. Chem. Phys.*, 1981, **74**, 3316–3319.
49 J. E. Jones, *Proc. R. Soc. London, Ser. A*, 1924, **106**, 463–477.