Journal Name

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

Valence, loop formation and universality in self-assembling patchy particles

Debra J. Audus,*[a] Francis W. Starr,[b] and Jack F. Douglas[c]

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

Patchy particles have emerged as an attractive model for phase separation and self-assembly in globular proteins solutions, colloidal patchy particles, and molecular fluids where directional interactions are operative. In our previous work, we extensively explored the coupling of directional and isotropic interactions on both the phase separation and self-assembly in a system of patchy particles with five spots. Here, we extend this work to consider different patch valences and isotropic interaction strengths with an emphasis on self-assembly. Although the location of self-assembly transition lines in the temperature-density plane depend on a number of parameters, we find universal behavior of cluster size that is dependent only on the probability of a spot being bound, the patch valence, and the density. Using these principles, we quantify both the mass distribution and the shape for all clusters, as well as clusters containing loops. Following the logical implications of these results, combined with a simplified version of a mean-field theory that incorporates Flory-Stockmayer theory, we find a universal curve for the temperature dependence of cluster mass and a universal curve for the fraction of clusters that contain loops. As the curves are dependent on the patchy valence, such results provide a method for parameterizing patchy particles models using experimental data.

1 Introduction

Short-ranged, directional interactions play a crucial role in controlling the structure and assembly of a variety of important systems, ranging from molecular fluids to small globular proteins to colloidal solutions.[1–4] Modeling these systems as patchy particles has received extensive attention, due to the ability of these simple models to capture non-trivial self-assembly and phase separation phenomena of real complex fluids such as reentrant phase behavior.[5–7] The popularity of patchy particles, composed of spheres with patches on the surface that introduce directional interactions, can also be attributed to the ability of analytic expressions such as Wertheim's thermodynamic perturbation theory[8–10] and extensions[11] to accurately describe phase separation and self-assembly, thereby reducing the reliance on computationally expensive simulations.

In the context of small globular proteins, patchy particles represent a useful coarse-grained model,[12–21] as the patches can represent charged or hydrophobic residues displayed on the surface. For example, Vlachy and coworkers used Wertheim theory to reproduce the liquid-liquid phase boundaries of both lysozyme and γ-crystalline using the number of patches as a fitting parameter.[20] Following their initial work, they considered a mixture of γ-crystallines, modeling each protein as a dumbbell of patchy particles.[21] While Vlachy and coworkers used models that considered only a hard sphere isotropic interaction, Kumar and coworkers added a short-range, isotropic attractive interaction to also account for the liquid-liquid phase coexistence of both lysozyme and γ-crystalline, using a combination of Monte Carlo simulations and theory.[18] For more complicated systems such as salt dependent reentrant phase coexistence of human serum albumin with a multivalent cation, an ion-activated patchy particle model combined with Wertheim theory qualitatively reproduces experiments.[17] The success of patchy particle models, and variants thereof, highlight their power to describe complicated protein solutions at various concentrations, including non-physiological ones. However, much of this work focused on phase separation rather than on self-assembly, which is also known to occur experimentally as a distinct process.[22]

From a practical standpoint, patchy particles are also of interest to colloidal scientists due to their potential for applications, such as electronics and drug delivery.[23–25] Numerous synthesis techniques[23,26,27] are available, many of which allow for control of patch size, strength and directionality, including the recent techniques of block copolymer assembly[28] and DNA-coated colloids.[29]

[a] Materials Science and Engineering Division, National Institute of Standards and Technology, Gaithersburg, MD 20899 USA. E-mail: debra.audus@nist.gov
[b] Physics Department, Wesleyan University, Middletown, CT 06459 USA.
[c] Materials Science and Engineering Division, National Institute of Standards and Technology, Gaithersburg, MD 20899 USA. E-mail: jack.douglas@nist.gov

Due to the widespread adoption of patchy particles models and the applicability of Wertheim theory, patchy-particle models have been extensively studied via theory and computation.[1–4] In particular, Sciortino and coworkers have made numerous contributions,[30,31] including finding that increasing number of patches widens the phase coexistence boundaries[5] and demonstrating the applicability of Flory-Stockmayer (FS) theory.[6] Telo da Gama and coworkers have explored loop formation in the case of patchy particles with two types of patches,[7] while Frenkel and coworkers determined the effect of adding an isotropic attractive interaction on the gas-liquid critical point.[32] While much progress has been made in understanding patchy particles, most efforts have not examined the universal nature of self-assembly, particularly of loop formation, that occurs in regions of the phase diagram without phase coexistence. Loop formation is of particular interest, as looping appears to lengthen the cluster lifetimes and to result in more compact cluster shapes. Although beyond the scope of this paper, loops are highly relevant to the properties of a self-assembled network beyond the point of percolation. Additionally, from the standpoint of predicting patchy particle self-assembly, the highly successful FS theory assumes that no loops are present in its derivation. This leads to the question of whether FS theory can still accurately predict cluster size distributions in cases where loops form.

In our prior work,[33] we focused on the introduction of isotropic interactions to examine the interplay between phase separation and self-assembly. We showed that Werthiem theory accurately describes phase separation in the absence of isotropic, attractive interactions, and, through the introduction of an analytic adjustment, can also be used for the case the where isotropic attractions are included. We also thoroughly explored the self-assembly of 5 spot patchy particles, showing how Werthiem theory can be combined with FS to predict the universal dependence of cluster mass on temperature. Such a result suggests that the theory can be used to more rigorously parametrize patchy particle models based on experimental measurements. Here, we extend this work[33] to show that the framework developed there can be used for variable patch numbers and isotropic attraction strengths, and that the validity holds beyond expected limits. We also quantify the resulting self-assembled clusters, exploring the formation of loops and associated universal behavior.

The paper is organized as follows. In Sec. 2, we introduce our model for patchy particles, present a variant of Wertheim theory and detail simulation techniques. In Sec. 3, we focus on the self-assembly of the patchy particles, first by identifying the region of self-assembly and then quantifying the self-assembled clusters including their size, shape and the presence of loops. An analysis of this kind is provided for different patch valences including 2, 3, 5, and 7 spots. We show that the basic concepts from our earlier work hold for the other patch valences and additionally use insights to understand the formation of loops. Finally, in Sec. 4, we summarize the key results.

2 Methods

2.1 Patchy Particle Model

We consider patchy particles with four different numbers of spots, or valances: 2, 3, 5 and 7, with the placement of these patches shown in Fig. 1. Two spots are chosen as their behavior is fundamentally different than patchy particles with larger valances, as demonstrated below. Odd values of spot numbers for 3 and greater are chosen to cover a wider range of valances, while avoiding crystalization that occurs for 6 spots as 6 spots correspond to a perfect cubic lattice using the same scheme for choosing spot locations.[18]

These patchy particles interact through two different potentials. The first is an isotropic potential u_{i} that is defined as a function of the distance between the centers of the red spheres, r_{i}.

$$u_{\mathrm{i}}(r_{\mathrm{i}}) = \begin{cases} \infty & r_{\mathrm{i}} \leq \sigma \\ \varepsilon_{\mathrm{i}} & \sigma < r_{\mathrm{i}} \leq \lambda\sigma \\ 0 & r_{\mathrm{i}} > \lambda\sigma \end{cases} \tag{1}$$

Here σ is the diameter of the red sphere, λ is the range of the isotropic square well potential and ε_{i} is the depth of the isotropic square well potential. $\lambda = 1.15$ and various values of ε_{i} are chosen to correspond to our prior work.[33] The second potential u_{p} is a square well potential between patches; thus for each pair of patchy particles, there are multiple patch interactions.

$$u_{\mathrm{p}}(r_{\mathrm{p}}) = \begin{cases} \varepsilon_{\mathrm{p}} & r_{\mathrm{p}} \leq \delta\sigma \\ 0 & r_{\mathrm{p}} > \delta\sigma \end{cases} \tag{2}$$

r_{p} is the distance between blue spots, and ε_{p} is the depth of the patchy square well potential. δ is the size of the spots and is chosen to be $(\sqrt{5-2\sqrt{3}}-1)/2 \approx 0.119$, which is the largest patch size to enforce the constraint that three patches cannot mutually interact to form a single "bond." Later, we relax this constraint to test theories beyond their limit of validity, doubling δ.

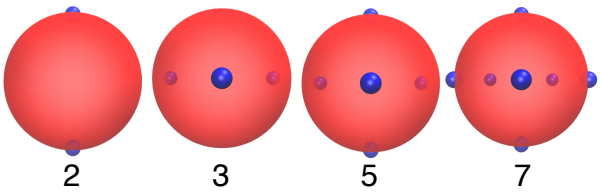

Fig. 1 Patchy particles with number and location of spots shown. Three and five spots have spots located at the points of an equilateral triangle around the equator, while seven spots has spots located at the points of an equilateral pentagon around the equator. This image was generated using VMD software.[34]

2.2 Monte Carlo Simulations

In order to simulate the patchy particles, off-lattice, Metropolis Monte Carlo simulations were run with a 50 % probability for both single particle displacement and single particle rotation. Both maximum displacement and maximum angle were chosen such that acceptance of a trial was roughly 50 %. In or-

der to ensure equilibrium conditions, simulations with densities $\rho \equiv N/V < 0.8\sigma^{-3}$ were initialized with a grand canonical simulation until the target density was reached. For $\rho \geq 0.8\sigma^{-3}$, the grand canonical simulation was run only once to generate an initial configuration that was then used for all temperatures and interaction strengths, as the procedure for high densities was prohibitive. Using the initial configuration, the simulations were run until the desired acceptance ratios were reached. Subsequently, the simulations were run for 5×10^8 timesteps for equilibration and 9.5×10^9 timesteps for production. All simulations were run with a periodic cubic simulation box with side lengths of 20σ. Our previous work showed that box lengths of 10σ were sufficient for 5 spots,[33] but we found that the larger box sizes were required for smaller numbers of spots, specifically the 2 spot case, in order to capture highly extended clusters for large cluster sizes.

2.3 Renormalized Mean-Field Theory

A renormalized mean-field theory is used to test the validity of theory in describing our Monte Carlo simulations. Specifically, we use a renormalized version of a statistical associating field theory for variable range potentials (SAFT-VR).[11] We have previously described the details of the renormalized mean-field theory (RMFT),[33] but highlight some of the key aspects and dependencies here. Specifically, the Helmholtz free energy normalized by the number of particles is given by

$$f = f_{\mathrm{id}} + f_{\mathrm{i}} + f_{\mathrm{p}}, \tag{3}$$

where the subscripts id, i and p correspond to the ideal, isotropic and patchy contributions, respectively.

The ideal contribution depends only on density ρ, whereas the isotropic contribution, derived using an inverse temperature expansion[35,36] and the Carahan and Starling hard sphere contribution,[37] is dependent on ρ, λ and ε_{i}. The patchy contribution is derived using Wertheim's thermodynamic perturbation theory[8–10] and results in

$$\beta f_{\mathrm{p}} = s\left(\ln X - \frac{X}{2}\right) + \frac{s}{2}. \tag{4}$$

β is the inverse of $k_B T$, s is the number of spots, and X is the fraction of patches that are non-bonded. The value of X can be computed analytically as

$$X = \frac{2}{1+\sqrt{1+4s\rho\Delta}} \tag{5}$$

where Δ is the patch interaction strength and depends on ρ, ε_{p}, δ, ε_{i}, and λ. See Ref. 33 for details.

The resulting mean-field theory does not predict the correct theta temperature T_Θ, or equivalently Boyle temperature, defined as the temperature at with the second virial coefficient is zero, *i.e.*, $B_2(T_\Theta) = 0$. To correct this deficiency, we define a renormalized $\varepsilon_{\mathrm{i}}^{\mathrm{re}}$ that satisfies $T_\Theta^{\mathrm{MFT}}(\varepsilon_{\mathrm{i}}^{\mathrm{re}}) = T_\Theta(\varepsilon_{\mathrm{i}})$. As T_Θ can be computed exactly for our system, this represents a straightforward way to ensure the exact T_Θ is satisfied for any ε_{i}. Note that no renormalization procedure is necessary for $\varepsilon_{\mathrm{i}} = 0$, as the above expression is automatically satisfied.

For the remainder of the paper, the energy and temperature scales will be defined relative to ε_{p}, while the length scales will be defined relative to σ, as in our prior work. Mass of a particle will be set as 1 such that cluster size and cluster mass are equivalent.

2.4 Estimation of phase boundaries

As Monte Carlo simulations of the phase boundaries are computationally expensive and only estimates of phase boundaries are necessary for our goals, we make use of the RMFT to estimate phase boundaries. In our previous work,[33] we found that the estimates of the critical point from RMFT were significantly different from those determined using Monte Carlo. To address this, we proposed a theory for computing the ratio of the critical temperature between the Monte Carlo simulations and RMFT ($T_{\mathrm{c}}^{\mathrm{sim}}/T_{\mathrm{c}}^{\mathrm{RMFT}}$). The theory only requires knowledge of a single Monte Carlo simulation of isotropic spheres without patches, and thus can be used to estimate $T_{\mathrm{c}}^{\mathrm{sim}}$ for patchy particles (see Fig. 3b in Ref. 33). Thus, for estimates of the phase boundary, we simply multiplied the temperature by the $T_{\mathrm{c}}^{\mathrm{sim}}/T_{\mathrm{c}}^{\mathrm{RMFT}}$ to produce reliable estimates. Note that this ratio is dependent on both ε_{i} and the number of spots s.

3 Results & discussion

3.1 Self-assembly transition lines

Before quantifying the self-assembled structures that result from patchy particles with different valences, or equivalently, number of patches, it is useful to map out the regions in the temperature-density plane where self-assembly occurs. The observed self-assembly is a thermoreversible or "rounded" transition,[38,39] rather than phase separation. To quantify this process, we make use of two metrics to define self-assembly transition lines. The first metric is based on the extent of clustering, Φ, and the second metric is based on the probability of percolation, p_{perc}. For both of these metrics, it is essential to define when a particle is in a cluster. As the patch interactions are of the form of a square-well potential, clusters are precisely defined through bonds where the patch potential is non-zero. Since liquid-liquid phase separation occurs at low temperatures, we also estimate the phase boundaries as a reference point, the details of which can be found in Sec 2.4. As can be seen in Fig. 2, the phase boundaries migrate to higher temperatures as the valance increases. This result is unsurprising given that the systems with higher valances are more attractive to one another, promoting phase separation.

The first metric for assembly captures the extent of clustering, Φ, which is defined as the fraction of particles that are in a cluster as opposed to being unbound, or equivalently, monomers. At sufficiently high temperatures, $\Phi = 0$, and all of the particles are unbound; at the other extreme of low temperature, assuming phase separation has not occurred, Φ approaches 1, and all particles are in clusters. For evident reasons, Φ is an order parameter for self-assembly. However, these clusters can be distant, and their respective sizes are not captured directly by Φ. As commonly employed in self-assembling systems, the midpoint of assembly, $\Phi = 1/2$, is used to characterize to thermodynamic location of assembly. Thus, the metric, T_Φ, is determined by computing the temperature at which $\Phi = 1/2$ for different densities as shown in

Fig. 2 for both the simulations and the RMFT. RMFT accurately estimates the transition.

The second metric for assembly quantifies the emergence of clusters that percolate the system. From the perspective of a simulation, the probability of percolation, p_{perc}, can be defined as the fraction of simulation snapshots that contain at least one percolating cluster, that is clusters that wrap around the periodic boundary conditions to bond to themselves. At high temperatures, there are no clusters; thus, the probability of percolation is zero. As temperature is decreased this probability increases. In the thermodynamic limit, p_{perc} as a function of temperature should be a Heaviside step function.[40] However, due to finite size effects, some rounding of the transition occurs. Thus, as an estimate of where the step occurs, we have used $p_{\text{perc}} = 1/2$ as shown in Fig. 2. We have previously shown that although the degree of rounding is sensitive to finite size effects, the second metric, T_p, is not.[33]

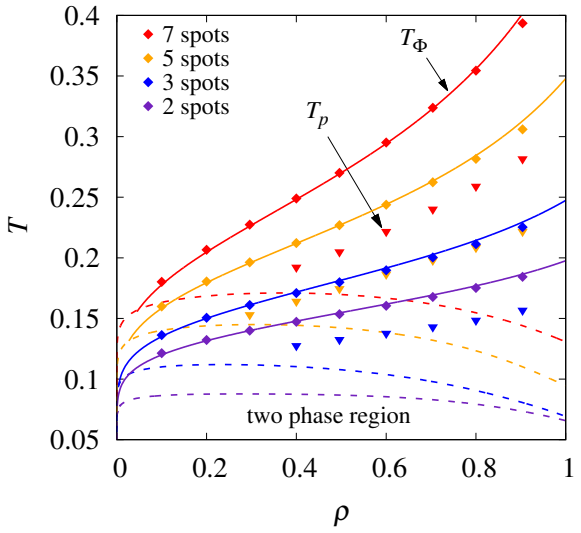

Fig. 2 Regions of self-assembly as defined using two metrics, T_Φ (simulations: diamonds and RMFT: solid lines) and T_p (simulations: triangles). For reference, estimates of the phase separation boundaries (dashed lines) are also shown. $\varepsilon_i = 0.1$.

Together, these transition lines, T_Φ and T_p indicate the condition of the system: at high temperatures, all of the particles are by themselves; as the temperature decreases, the extent of clustering continues to increases as it passes through T_Φ; further decrease of temperature leads to percolation; eventually, the two phase region is reached. We find that the region of self-assembly, roughly defined as the area under the T_Φ curve is larger for larger valences, which is consistent with intuition as a larger number of patches promotes self-assembly.

While Fig. 2 only considers the case where the isotropic attractive interaction, ε_i, is 0.1, the results are qualitatively similarly for other values of ε_i: 0, 0.2, 0.3 and 0.4. As shown in our previous paper, increasing ε_i promotes phase separation and reduces the one phase region of self-assembly.[33]

3.2 Mean cluster size & cluster shape

With the map of self-assembly, we next examine the cluster size distributions. In Fig. 3, we plot the probability of finding a cluster of size (mass) M. The open symbols correspond to all clusters, while the closed symbols correspond to only the clusters that contain loops. For the entire population, we find that the size distribution can be described remarkably well by the mean-field FS theory,[41,42] a model that assumes loops are not present. From the FS theory, the probability distribution can be described as

$$p(M) = \frac{2sX^{2+M(s-2)}(1-X)^{M-1}(sM-M)!}{(sX+2-s)(sM-2M+2)!M!} \qquad (6)$$

where X is the probability of a spot being unbound and s is the number of spots, or equivalently valence. Since X is an average quantity, it is equivalent to the fraction of spots that are unbound. It can equivalently be estimated directly from simulation as $\Phi = 1 - X^s$. Physically, this equation can be understood by considering that the probability for an unbonded particle is simply the probability that all of its patches are unbound, X^s, while the fraction of particles in a cluster is one minus that quantity. Since, both Φ and s are known, X can be estimated and $p(M)$ can be determined, which results in the solid line through the open symbols. The results for the 2 spot case are consistent with work by Sciortino and coworkers.[30] In the limit of large M, FS theory must break down in part due to the higher fraction of looped clusters at large M. It has been previously shown that patchy particle self-assembly follows a power law for large M and that the exponent is not consistent with FS theory.[33] Nonetheless, the theory does capture the size distribution for higher probabilities, corresponding to smaller clusters, rather well.

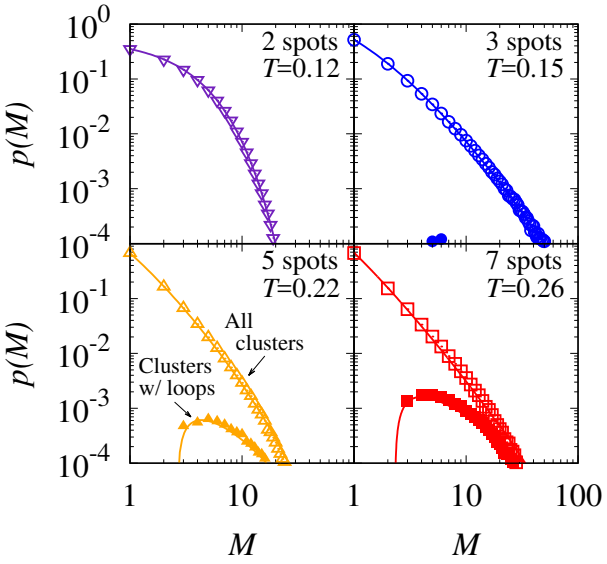

Fig. 3 Cluster size distributions for all clusters (open symbols) and those that contain loops (closed symbols). Lines correspond to FS theory for all clusters and theory as described in the text for loop containing clusters. All data is for $\rho = 0.6$.

We also consider the size distribution for clusters that contain loops. In the case of the 2 spots particles, no rings are observed, although rings are not prohibited as even with geometric constraints as few as 8 particles are needed to form a ring. The reason for this absence of rings will be explored further when we consider the shape of the clusters. In the case of 3 spots, loops can form, but they tend to form with extremely low probabilities that make it hard to acquire accurate statistics. For the 5 and 7 spot cases, the relative probability of forming clusters with loops for a given cluster size is significant larger than that of 3 spots as can be seen in Fig. 3.

Interesting, the probability of finding a loop-containing cluster of a given size can also be described analytically for the 5 and 7 spot cases using an empirical expression. In particular, the probability can be described by

$$p_{\text{loop}}(M) = \begin{cases} \frac{(1-X)(M-\alpha)}{\beta\rho} p(M) & M > 2 \\ 0 & M \leq 2. \end{cases} \quad (7)$$

This equation involves two additional parameters, α and β. Note that if $p_{\text{loop}}(M)/p(M)$ is plotted as a function of M, this results in a line with a x intercept of α and a slope of $(1-X)/(\beta\rho)$. As clusters with two particles ($M = 2$) cannot form loops, while clusters with 3 particles ($M = 3$) can form loops, the x intercept, or equivalently α, must be between 2 and 3. Both α and β are only dependent on the number of spots and not the temperature, density or isotropic interaction strength, parameters that are implicitly captured in X and ρ in Eq. 7. Although this equation is derived empirically, the predicted trends are consistent with physical intuition. Larger clusters, corresponding to large M, have larger fractions of loops as there are more particles present in the cluster to form loops. Loops are also more likely to form when probability of bonding $(1-X)$ is high, a situation that occurs at low temperatures and high densities; it is also not unreasonable to expect loop formation might be linear with respect to the probability of bonding as loops are formed by one or more additional bonds within a cluster compared to loop-less clusters. Finally, the ρ dependence is more complicated as it is dependent both implicitly in $1-X$ and explicitly in ρ in Eq. 7. For a fixed temperature, and isotropic interaction strength, we find that the ρ dependence is dominated by the $1-X$ in the numerator rather than the ρ in the denominator such that the fraction of loops increases with ρ, an expected result as higher densities will force clusters to be more compact.

As shown below, the existence of such a universal distribution implies a master curve for the fraction of clusters in the system multiplied by the density. Using this master curve, we fit both of the unknown parameters finding that $\alpha = 2.6$ and $\beta = 22$ for 5 spots, while $\alpha = 2.3$ and $\beta = 7.5$ for 7 spots. Note that α is between 2 and 3, as expected. Although only one probability distribution is shown in Fig. 3 for a given number of spots, this relationship holds for other temperatures, densities and isotropic interaction strengths. At some point the distribution in Eq. 7 must break down as it predicts that the fraction of loops increases linearly with cluster size, a quantity that is bounded to be less than or equal to one. This erroneous prediction could be corrected by using a piecewise function that sets $p_{\text{loop}}(M) = p(M)$ when $p_{\text{loop}}(M)/p(M) > 1$, but as p_{loop} is small for large M, such a contribution can be ignored. As this result can be used for at least 5 spots and larger, it provides an interesting way of predicting the full distributions for all clusters and clusters with loops using only theory, as the RMFT can be used to calculate Φ, which can then be used to determine X and the distributions, assuming α and β are known for a given number of spots (and likely spot size).

Having quantified the size distributions, we consider the shape of the clusters, both with and without loops. In Fig. 4a, we plot the radius of gyration of the clusters as a function of cluster size, which determines the mass scaling exponent ν, $\langle R_g^2 \rangle^{1/2} \sim M^{\nu}$. For 3, 5 and 7 spots, we find that once the clusters get large enough, they follow a power law with $\nu = 1/2$. For 5 and 7 spots, this is true both for clusters that contain loops and those that do not. For 3 spots, the power law also holds for clusters without loops for large enough M. However, no definitive conclusion regarding loops can be formed as there are only $\langle R_g^2 \rangle^{1/2}$ values for $M < 10$. The power law $\nu = 1/2$ corresponds to a fractal dimension $d_f = 1/\nu = 2$ and is consistant with lattice animals,[43] but distinct from percolation clusters, which have a fractal dimension near 2.5.[40] The former is in the same universality class as swollen branched polymers, while the latter is in the same universality class as branched polymers in a theta solvent.[44] Also note that the fractal dimension predicted by the mean-field FS theory is 4,[40] which is rather different than the observed value of 2.

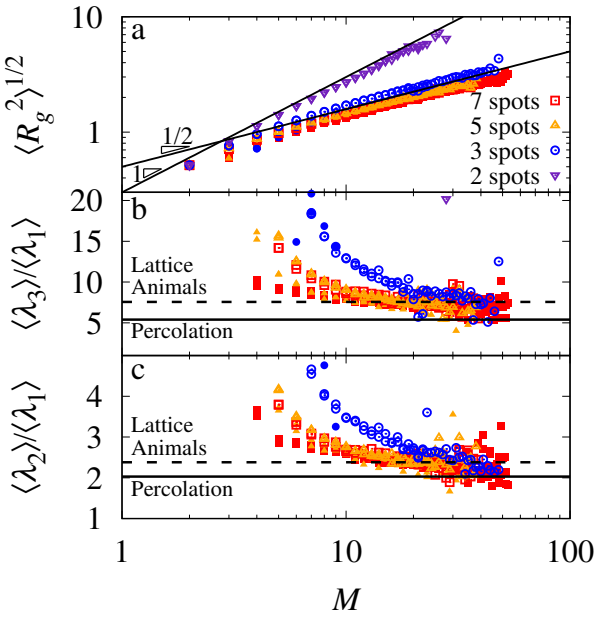

Fig. 4 Radius of gyration and ratios of eigenvalues of the gyration tensor for various cluster sizes and both those with (open) and without (solid) loops for selected conditions.

For the case of 2 spots, the radius of gyration also follows a power law, but with $\nu = 1$, rather than 1/2. This corresponds to a fractal dimension of one, and is consistent with a rod-like shape. The reason for this rod-like behavior is the constrained nature of the bonds. In particular, the particles are composed of

large spheres with much smaller patches. The maximum angle that two particles can bend from a straight line is 57°, and a large degree of bending is unlikely due to entropic considerations. This also explains why we observe no closed rings (loops) for the 2 spot case.

The shape of the clusters can be further explored by computing the eigenvalues of the radius of gyration tensor, λ_j, where $\lambda_1+\lambda_2+\lambda_3=R_g^2$ and $\lambda_3>\lambda_2>\lambda_1$. The ratio of these eigenvalues quantifies how anisotropic the particles are.[45] If all three eigenvalues are equal, then the object is roughly spherical and both ratios will be 1. Conversely, if two of the eigenvalues are 0, then the object is a rod and one ratio is infinity and the other ill-defined. As the 2 spot clusters are highly anisotropic, they, with the exception of a single point, are off the scale of the plot of the ratios in Fig. 4b and c.

For 3, 5 and 7 spots, as the clusters become large enough, the ratios λ_3/λ_1 and λ_2/λ_1 approach those expected for lattice animals, 7.56 and 2.03, respectiviely.[45] Lattice animals, as previously mentioned, are in the same universality class as branched polymers in a good solvent. For comparison, the limit for percolation clusters,[45] which are in the same universality class as branched polymers in a theta solvent is also plotted. As the number of spots increase, the large M limit is reached sooner. The limit is also reached sooner for clusters that contain loops, as those clusters tend to be more compact when M is small.

3.3 Universal metrics of cluster size

In our last paper,[33] we explored the possibility of a universal functional relationship to describe the variation of the mean cluster size $\langle M\rangle$ for the 5 spot case over a range of isotropic interaction strengths. Here we extend it to other valences and test the limits of the theory, demonstrating that the theory can hold even beyond its limits. Ultimately, we want to determine the average mass as a function of temperature, as this can be measured experimentally. This can be done by linking the average mass to X and, subsequently, X to temperature.

First, we consider the average cluster size. As we showed above the entire mass distribution can be described by the FS theory, which requires only X as an input. Thus, the average mass can also be computed as a function of the fraction of non-bonded patches X simply by computing the appropriate sum.

$$\langle M\rangle=\frac{2}{sX+2-s} \tag{8}$$

for $X\geq 1-1/(s-1)$. For $X<1-1/(s-1)$, FS predicts percolating clusters and the sum to compute $\langle M\rangle$ is not convergent. As X is not directly measured from simulation, we can convert X to Φ using $\Phi=1-X^s$ where s is the number of spots, or equivalently valence. The result can be found in Fig. 5 for different number of spots. We find excellent agreement. The gray lines correspond to the case beyond the limit of the validity of FS theory, that is, when the theory predicts percolation. The points plotted in Fig. 5 correspond to different number of spots, different temperatures, different densities, and different isotropic interaction strengths. The only constraint is that the temperature and density must be chosen such that it is above both the two phase region and percolation lines shown in Fig. 2. The fact that the simulation data continues to larger Φ than the range of validity of FS theory implies that FS underestimates the value of Φ necessary for percolation. The reason for this deviation is likely due to the fact that FS does not consider loops, which are clearly present in the simulations. Nonetheless, FS still reproduces essential trends well beyond its strict range of validity. In order to further test if the theory can be effective beyond its formal range of validity, we also consider a system of patchy particles containing 5 patches, but with patches that are large enough to break the assumption of the RMFT; specifically, we considered spots with a diameter of twice the size. These also follow the same trend suggesting that theory may still be useful for characterizing experimental systems even if the assumptions of the theory are not fully realized.

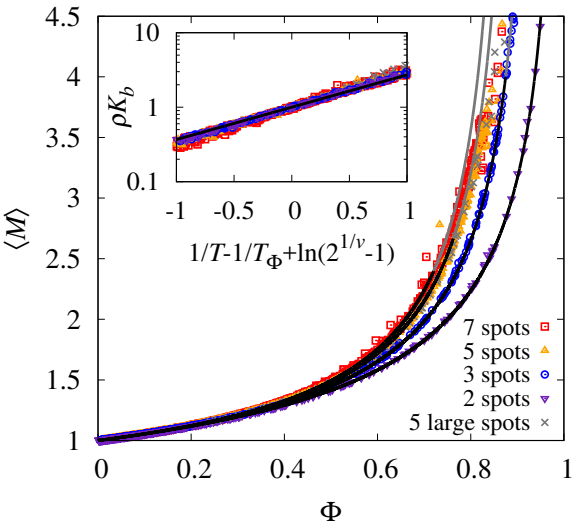

Fig. 5 Master curves for extent of clustering and ρK_b (inset). Points correspond to simulation data for different ρ (0.1 to 0.9), T (<0.4) and ε_i (0 to 0.4) above both the binodal and the percolation transition. Black and gray lines correspond to the theory within and beyond the limits of validity, respectively. Data for 5 large spots is beyond the validity of the theory.

Second, we consider linking the temperature to X, since temperature is the operational variable experimentally. This can be done by treating the formation of a bond as a reaction. In this case, the mass-action law gives

$$\frac{1-X}{X^2}=s\rho\Delta=\rho K_b, \tag{9}$$

where K_b is the reaction constant. Using the RMFT and simplifying assumptions as explained in Ref. 33,

$$\ln\left(\frac{1-X}{X^2}\right)=\frac{1}{T}-\frac{1}{T_\Phi}+\ln\left(2^{1/s}(2^{1/s}-1)\right). \tag{10}$$

Assuming a simple activated process,

$$\rho K_b=\exp\left(-\frac{\Delta F}{T}\right)=\exp\left(-\frac{\Delta U}{T}+\Delta S\right), \tag{11}$$

where ΔU and ΔS are the energy and entropy of activation, their values are

$$\Delta U = -1 \tag{12}$$

and

$$\Delta S = -\frac{1}{T_\Phi} + \ln\left(2^{1/s}(2^{1/s}-1)\right). \tag{13}$$

We now test the validity of these simplifying assumptions. From simulations, we can indirectly compute X via Φ; see inset in Fig. 5. The line is the prediction from theory, and the data for all the different number of spots follows the line closely. Deviations only occur for the largest number of spots, 7 spots, and these deviations are small. The minor discrepancy is likely due to the breakdown of the simplifying assumptions applied to the RMFT to yield eq. 10.

Using the results above, a master curve for the average mass as a function of the temperature can be generated by connecting two master curves above. Specifically, combining Eq. 8, Eq. 9 and Eq. 10 yields a universal curve as plotted in Fig. 6. We see excellent agreement both within and even beyond the limits of validity, which is not surprising given the performance of the two master curves in Fig. 5. To reemphasize the point, this plot is for all valences, temperatures, densities and interaction strengths such that the state point is in the one phase region above the percolation line. Although not perfect, this approximation does well even for the case of 5 large spots, where the theory is pushed beyond its range of assumptions.

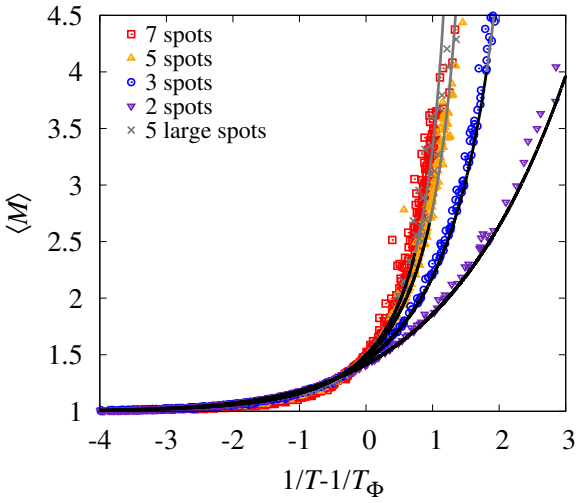

Fig. 6 Master curve for the average cluster size. Points correspond to simulation data for different ρ (0.1 to 0.9), T (<0.4) and ε_i (0 to 0.4) above both the binodal and the percolation transition. Black and gray lines correspond to the theory within and beyond the limits of validity, respectively. Data for 5 large spots is beyond the validity of the theory.

The master curve above has interesting experimental implications. It means that if one knows the average mass as a function of temperature for various densities, the data can be used to parameterize a patchy-particle model, extract T_Φ, the energy, and the entropy. In our prior paper, we suggested this could be done using light scattering following the approach of Burchard.[46,47] However, we have since found that the required assumptions are incompatible with our simulations, as scattering curves can be extracted from simulation and the predicted molecular mass following Burchard is not consistant with the directly measured mass. An attractive, alternative approach may be sedimentation velocity analytical ultracentrifugation (AUC)[48–51] experiments, which can determine the entire cluster size distribution. However, care must be taken to consider the time-scale of association relative to the time-scale of the experiment as the clusters are separated by their sedimentation rates, which are dependent on the cluster size and assumed to be fixed during the the course of the experiment.[51]

Note that as the entire FS distribution is known, the appropriate master curve can be generated for any type of average mass. For example, the expression linking X to the mass average mass, M_w, as opposed to the number average mass, $\langle M\rangle$, can also be computed.

$$\ln\frac{1-X}{X^2} = \ln\frac{2s\langle M\rangle(\langle M\rangle-1)}{((s-2)\langle M\rangle+2)^2} \tag{14}$$

$$\ln\frac{1-X}{X^2} = \ln\frac{(M_w-1)(M_w(s-1)+1)}{(M_w(s-2)+2)^2} \tag{15}$$

This can be linked using Eq. 10 to yield the equivalent of Fig. 6. Additionally, the x axis of Fig. 6 can also be manipulated into an equivalent reduced form $[(T-T_\Phi)/T](1/T_\Phi)$, if a reduced representation that is similar to those used in critical phenomenon is desired.

3.4 Universal metric for loop containing clusters

For the universal curves above, we considered all clusters both those containing loops and those not containing loops. As mentioned earlier, the mean-field theory of FS neglects such loops. Nonetheless, a distinct universal curve can be generated for the loop containing clusters; this logically follows, as we empirically know the distribution for the clusters with loops as a function of X (see Eq. 7), thus, the fraction of clusters that contain loops as a function of X can be computed. f_{loop} is the number of clusters that contain loops divided by the total number of clusters,

$$f_{\text{loop}} = \frac{\sum_{i=3}^{\infty} p_{\text{loop}}(M)}{\sum_{i=1}^{\infty} p(M)}. \tag{16}$$

Analytically, performing these sums results in the empirical expression:

$$f_{\text{loop}}\rho = \frac{(1-X)}{\beta}\left(\frac{2+(2\alpha-2)X^s+(\alpha-2)sX^{2s-2}(1-X)}{2-s(1-X)} - \alpha\right), \tag{17}$$

which is tested in Fig. 7 for 5 and 7 spots. The gray portions of the lines are beyond the limit of assumptions where FS theory predicts percolation has already occurred. We find excellent agreement in all cases over nearly six orders of magnitude, but only within the limits of FS theory.

The primary source for the deviation beyond the limits of FS is likely from Eq. 7, the basis for Eq. 17 rather than the inherent assumptions in FS. As discussed previously, Eq. 7 must break down for large M as it predicts $p_{loop}(M)/p(M)$ increases linearly with M whereas physically, it must be bound to 1 as the number

of clusters with loops cannot be greater than the total number of clusters. This could be fixed by imposing $p_{loop}(M)/p(M) = 1$ for cases where Eq. 7 currently predicts $p_{loop}(M)/p(M) > 1$; however, the equivalent of Eq. 17 would no longer be analytic, reducing the practical utility of the theory. Upon closer examination, $p_{loop}(M)/p(M) \sim (1-X)$, so this issue is more likely to be problematic when $(1-X)$, or equivalently, the probability of a spot being bound is large. This is exactly the scenario under which Eq. 17 fails. It is also intuitive as the larger the probability of bonding, the larger the $p_{loop}(M)$ and thus the larger the erroneous contribution to the sum in Eq. 16. Nonetheless, under most conditions realized in practice, the error from Eq. 7 is negligible and Eq. 17 retains its predictive power. Thus, if an experimental system is found to have 5 or more spots, and the fitting parameters α and β are known, the fraction of loops can be computed, as well as its full distribution via Eq. 7. Also, f_{loop} can be determined as a function of temperature by combining Eqs. 10 and 17

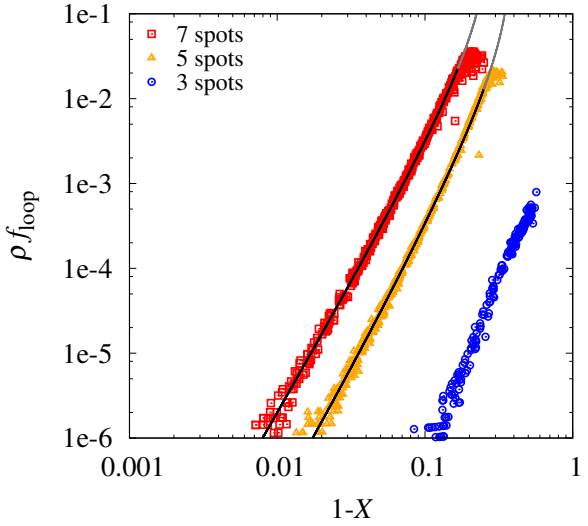

Fig. 7 Master curves for the density times fraction of clusters that contain a loop. Points correspond to simulation data for different ρ (0.1 to 0.9), T (< 0.4) and ε_i (0 to 0.4) above both the binodal and the percolation transition. Black and gray lines correspond to the theory within and beyond the limits of validity, respectively.

For the case of 3 spots, a master curve is also found from simulation. However, a similar analytic expression cannot be computed, as the expression in Eq. 17 does not hold for 3 spots. As the size of the clusters increases, the fraction of loops does not approach one, as in the case of 5 and 7 spots. From Fig. 4b and c, the 3 spot particles must form branches. Thus, many of the spots are used in forming branches, and extra spots for forming loops are limited. Therefore, there is not the same excess of spots as in the 5 and 7 spot cases, which makes the 3 spots qualitatively different, despite similar looking functional dependence in Fig. 7.

4 Conclusions

Using two metrics to define the degree of association, the extent of clustering and probability of percolation, we compute transition lines that can be used to quantify the regions of the temperature-density plane where self-assembly occurs. Although these self-assembly transition lines do not represent phase transitions, they do quantify the change in the state of self-assembly. Not surprisingly, we find that the region of self-assembly increases with increasing patch valence. Having defined the region of self-assembly above both phase coexistence and percolation, we quantify the resulting clusters. We first consider the mass distribution and show that the Flory-Stockmayer (FS) theory, an intrinsically mean-field model that neglects loops and breaks down for large M, accurately describes the distribution for small M, despite its deficiencies. If clusters with loops are considered, we find an empirical expression with two fitting parameters, which are only dependent on the valence, not the temperature, pressure, density, or isotropic interaction strength. This expression holds for both the 5 spot and 7 spot systems where the number of patches is in excess. Loops in the 2 spot system are not found due to the rod-like nature of clusters in this system. In addition to the mass distribution, we quantify the cluster shape finding that 3, 5 and 7 spot systems behave as lattice animals, and thus are in the same universality class as swollen, branched polymers, while the 2 spot system behaves as rods having a fractal dimension of one and a highly anisotropic shape.

Having quantified the properties of the clusters, we turn to the implications of the mass distributions. As implied by FS theory, the average mass must only depend on the probability of bonding, or equivalently the extent of clustering. This results in a universal curve that depends only on valence for the region of self-assembly above percolation and phase separation. Additionally, FS provides a description of this universal curve without any fitting parameters. Furthermore, the temperature is linked to the extent of clustering via the RMFT and a simplifying assumption. Although this assumption cannot be justified *a priori*, it does in fact hold for the simulation data, providing the necessary link, as well as a route to identify the energy and entropy of patch binding. Combined with FS theory, it produces a universal curve for average mass as a function of inverse temperature with no fitting parameters and only knowledge of the patch valence. From an experimental point of view, this provides a useful method for parametrizing patchy particle models using experimental data, specifically in the case of small globular proteins. Additionally, we consider the implications of the mass distribution for loops. After performing the appropriate sums, we also find universal behavior for the fraction of clusters that contain loops times the density as a function of the probability of patch bonding for cases where the number of patches is in excess. In this case, only two fitting parameters are required and the result is consistent with that for the full mass distribution. These results have significant implications for experimental quantification of systems that can be described by the model of patchy particles.

Conflict of interest

There are no conflicts to declare.

Acknowledgements

FWS acknowledges support from NIST award 70NANB15H282. Official contribution of the U.S. National Institute of Standards

and Technology - Not subject to copyright in the United States.

References

1 E. Bianchi, R. Blaak and C. N. Likos, *Physical Chemistry Chemical Physics*, 2011, **13**, 6397–410.

2 E. Bianchi, B. Capone, I. Coluzza and L. Rovigatti, *Limiting the valence: advancements and new perspectives on patchy colloids, soft functionalized nanoparticles and biomolecules*, 2017, arXiv:1705.04383v1 [cond-mat.mtrl-sci]. arXiv.org e-Print archive. https://arxiv.org/abs/1705.04383v1 (accessed Sept. 2017).

3 C. L. Porter and J. C. Crocker, *Current Opinion in Colloid and Interface Science*, 2017, **30**, 34–44.

4 E. Duguet, C. Hubert, C. Chomette, A. Perro and S. Ravaine, *Comptes Rendus Chimie*, 2015, **19**, 173–182.

5 E. Bianchi, J. Largo, P. Tartaglia, E. Zaccarelli and F. Sciortino, *Physical Review Letters*, 2006, **97**, 168301.

6 E. Bianchi, P. Tartaglia, E. La Nave and F. Sciortino, *The Journal of Physical Chemistry B*, 2007, **111**, 11765–11769.

7 J. M. Tavares, N. G. Almarza and M. M. Telo da Gama, *Soft Matter*, 2015, **11**, 5828–5838.

8 M. S. Wertheim, *Journal of Statistical Physics*, 1984, **35**, 19–34.

9 M. S. Wertheim, *Journal of Statistical Physics*, 1984, **35**, 35–47.

10 M. S. Wertheim, *Journal of Statistical Physics*, 1986, **42**, 459–476.

11 A. Gil-Villegas, A. Galindo, P. J. Whitehead, S. J. Mills, G. Jackson and A. N. Burgess, *The Journal of Chemical Physics*, 1997, **106**, 4168.

12 A. Lomakin, N. Asherie and G. B. Benedek, *Proceedings of the National Academy of Sciences of the United States of America*, 1999, **96**, 9465–8.

13 J. J. McManus, P. Charbonneau, E. Zaccarelli and N. Asherie, *Current Opinion in Colloid and Interface Science*, 2016, **22**, 73–79.

14 R. P. Sear, *The Journal of Chemical Physics*, 1999, **111**, 4800.

15 M. Hloucha, J. Lodge, A. Lenhoff and S. Sandler, *Journal of Crystal Growth*, 2001, **232**, 195–203.

16 J. K. Cheung, V. K. Shen, J. R. Errington and T. M. Truskett, *Biophysical Journal*, 2007, **92**, 4316–24.

17 F. Roosen-Runge, F. Zhang, F. Schreiber and R. Roth, *Scientific Reports*, 2014, **4**, 7016.

18 H. Liu, S. K. Kumar and F. Sciortino, *The Journal of Chemical Physics*, 2007, **127**, 084902.

19 C. Gögelein, G. Nägele, R. Tuinier, T. Gibaud, A. Stradner and P. Schurtenberger, *The Journal of Chemical Physics*, 2008, **129**, 085102.

20 M. Kastelic, Y. V. Kalyuzhnyi, B. Hribar-Lee, K. A. Dill and V. Vlachy, *Proceedings of the National Academy of Sciences of the United States of America*, 2015, **112**, 6766–70.

21 M. Kastelic, Y. V. Kalyuzhnyi and V. Vlachy, *Soft Matter*, 2016, **12**, 7289–7298.

22 A. Stradner, H. Sedgwick, F. Cardinaux, W. C. K. Poon, S. U. Egelhaaf and P. Schurtenberger, *Nature*, 2004, **432**, 492–5.

23 A. B. Pawar and I. Kretzschmar, *Macromolecular Rapid Communications*, 2010, **31**, 150–68.

24 L. Di Michele and E. Eiser, *Physical Chemistry Chemical Physics*, 2013, **15**, 3115–29.

25 K. J. Lee, J. Yoon and J. Lahann, *Current Opinion in Colloid & Interface Science*, 2011, **16**, 195–202.

26 G.-R. Yi, D. J. Pine and S. Sacanna, *Journal of Physics: Condensed Matter*, 2013, **25**, 193101.

27 J. Du and R. K. O'Reilly, *Chemical Society Reviews*, 2011, **40**, 2402–16.

28 A. H. Gröschel, A. Walther, T. I. Löbling, F. H. Schacher, H. Schmalz and A. H. E. Müller, *Nature*, 2013, **503**, 247–51.

29 Y. Wang, Y. Wang, D. R. Breed, V. N. Manoharan, L. Feng, A. D. Hollingsworth, M. Weck and D. J. Pine, *Nature*, 2012, **491**, 51–5.

30 F. Sciortino, E. Bianchi, J. F. Douglas and P. Tartaglia, *The Journal of Chemical Physics*, 2007, **126**, 194903.

31 J. Russo, P. Tartaglia and F. Sciortino, *The Journal of Chemical Physics*, 2009, **131**, 014504.

32 W. M. Jacobs, D. W. Oxtoby and D. Frenkel, *The Journal of Chemical Physics*, 2014, **140**, 204109.

33 D. J. Audus, F. W. Starr and J. F. Douglas, *The Journal of Chemical Physics*, 2016, **144**, 074901.

34 http://www.ks.uiuc.edu/Research/vmd/, VMD is developed with NIH support by the Theoretical and Computational Biophysics group at the Beckman Institute, University of Illinois at Urbana-Champaign.

35 J. A. Barker and D. Henderson, *The Journal of Chemical Physics*, 1967, **47**, 2856–2861.

36 J. A. Barker and D. Henderson, *The Journal of Chemical Physics*, 1967, **47**, 4714.

37 N. F. Carnahan and K. E. Starling, *The Journal of Chemical Physics*, 1969, **51**, 635.

38 J. Dudowicz, K. F. Freed and J. F. Douglas, *Journal of Chemical Physics*, 2000, **113**, 434–446.

39 V. Percec, G. Ungar and M. Peterca, *Science (New York, N.Y.)*, 2006, **313**, 55–56.

40 D. Stauffer and A. Aharony, *Introduction to Percolation Theory*, Taylor & Francis, 2nd edn., 1992.

41 P. Flory, *Principles of Polymer Chemistry*, Cornell University Press, 1953.

42 W. H. Stockmayer, *Journal of Chemical Physics*, 1943, **11**, 45.

43 G. Parisi and N. Sourlas, *Physical Review Letters*, 1981, **46**, 871–874.

44 J. F. Douglas, *Physical Review E*, 1996, **54**, 2677–2689.

45 M. L. Mansfield and J. F. Douglas, *The Journal of Chemical Physics*, 2013, **139**, 044901.

46 W. Burchard, *Makromolekulare Chemie. Macromolecular Symposia*, 1990, **39**, 179–195.

47 W. Burchard, P. Lang, L. Schulz and T. Coviello, *Makromolekulare Chemie. Macromolecular Symposia*, 1992, **58**, 21–37.

48 J. Lebowitz, M. S. Lewis and P. Schuck, *Protein Science*, 2002,

11, 2067–2079.

49 R. B. Gillis, G. G. Adams, B. Wolf, M. Berry, T. M. Besong, A. Corfield, S. M. Kök, R. Sidebottom, D. Lafond, A. J. Rowe and S. E. Harding, *Carbohydrate Polymers*, 2013, **93**, 178 – 183.

50 M. K. Phillips-Jones, R. Lithgo, V. Dinu, R. B. Gillis, J. E. Harding, G. G. Adams and S. E. Harding, *Scientific Reports*, 2017, **7**, 12697.

51 P. Schuck, H. Zhao, C. A. Brautigam and R. Ghirlando, *Basic Principles of Analytical Ultricentrifugation*, CRC Press, 2016.