PAPER

View Article Online
View Journal | View Issue

Macromolecular diffusion in crowded media beyond the hard-sphere model†

Pablo M. Blanco, [iD] *[ab] Josep Lluís Garcés,[c] Sergio Madurga*[ab] and Francesc Mas[ab]

The effect of macromolecular crowding on diffusion beyond the hard-core sphere model is studied. A new coarse-grained model is presented, the Chain Entanglement Softened Potential (CESP) model, which takes into account the macromolecular flexibility and chain entanglement. The CESP model uses a shoulder-shaped interaction potential that is implemented in the Brownian Dynamics (BD) computations. The interaction potential contains only one parameter associated with the chain entanglement energetic cost (U_r). The hydrodynamic interactions are included in the BD computations *via* Tokuyama mean-field equations. The model is used to analyze the diffusion of a streptavidin protein among different sized dextran obstacles. For this system, U_r is obtained by fitting the streptavidin experimental long-time diffusion coefficient D^{long} versus the macromolecular concentration for D50 (indicating their molecular weight in kg mol^{-1}) dextran obstacles. The obtained D^{long} values show better quantitative agreement with experiments than those obtained with hard-core spheres. Moreover, once parametrized, the CESP model is also able to quantitatively predict D^{long} and the anomalous exponent (α) for streptavidin diffusion among D10, D400 and D700 dextran obstacles. D^{long}, the short-time diffusion coefficient (D^{short}) and α are obtained from the BD simulations by using a new empirical expression, able to describe the full temporal evolution of the diffusion coefficient.

1 Introduction

Biological media are known to contain a high concentration of a wide variety of macromolecular species such as proteins, polysaccharides or nucleic acids. For instance, the weight fraction of protein is around 5% in lymph, 9% in blood plasma and 35% in hemolysate.[1,2] In the cellular cytosol, three-dimensional analysis of electron micrographs revealed a 20% volume fraction of fibrous supramolecular structures (*i.e.* F-actin, microtubules and intermediate filaments).[3] These conditions, known as "macromolecular crowding" involve non-specific interactions among macromolecular species due to the excluded volume, van der Waals, electrostatic and hydrodynamic interactions. Macromolecular crowding substantially alters the diffusion processes, conformational properties and reactivity of the macromolecular species.[4–8]

Several efforts are underway to properly understand the effect of crowding in macromolecular diffusion.[9] Recently, *in vivo* and

in vitro experimental studies have been carried out to study macromolecular diffusion in crowded media. In *in vivo* experiments, fluorescent tracer proteins are introduced into the cell by means of transfection, microinjection or recombinant expression.[10] Although this strategy allows the direct study of the macromolecular crowding, the intrinsic differences between the different intracellular microenvironments make the interpretation of the results difficult. *In vitro* experiments attempt to overcome this problem by reducing the complexity and heterogeneity of the media. The crowded environment is recreated using highly concentrated polymer solutions (usually dextran or ficoll).[11–13] The motion of a fluorescent protein is then studied using spectroscopic techniques,[14] mainly Fluorescence Correlation Spectroscopy (FCS)[11] and Fluorescence Recovery After Photobleaching (FRAP).[12,13]

In order to interpret the experimental results, an increasing number of computational studies have been performed[15] that provide a highly controlled environment. The comparison of the computational results with the experimentally observed results facilitates the quantification of the factors governing macromolecular diffusion in crowded media. Different computational approaches have been applied ranging from on-lattice Monte Carlo simulations[16,17] and off-lattice Brownian Dynamics (BD)[18–25] to Molecular Dynamics simulations.[23,26,27] In implicit solvent simulations like BD, the Hydrodynamic Interactions (HIs) of the macromolecules have been found to be crucial to properly describe their motion in crowded media.[28]

[a] Department of Material Science and Physical Chemistry, Barcelona University, 08028 Barcelona, Spain. E-mail: fmas@ub.edu, pmblanco@ub.edu, s.madurga@ub.edu

[b] Institute of Theoretical and Computational Chemistry (IQTC), Barcelona University, 08028 Barcelona, Spain

[c] Department of Chemistry, University of Lleida (UdL), 25003 Lleida, Spain. E-mail: jlgarces@quimica.udl.cat

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c8sm00201k

Macromolecular diffusion is known to become sub-diffusive in crowded media. The time-dependent diffusion coefficient (D) of the macromolecule is defined as $D \equiv \langle r^2 \rangle / (2dt)$, where d is the topological dimension of the system and $\langle r^2 \rangle$ is the mean square displacement of the particles. In crowded media, D undergoes a transition from an initial value at short times (D^{short}) to an asymptotic final value at long times (D^{long}). During the transition between both regimes, macromolecular motion shows a non-linear dependence on time (usually called anomalous diffusion):[16,29]

$$\langle r^2 \rangle = (2d)\Gamma_\alpha t^\alpha \qquad (1)$$

where α is the anomalous diffusion exponent and Γ_α is a generalized transport coefficient.

In order to study macromolecular diffusion in the long-time regime, it is necessary to reach μs time scales. However, the large number of atoms present in a highly concentrated macromolecular solution causes atomistic-detailed computations to be limited to smaller time scales. As a consequence, in order to reach larger time scales, coarse grained approaches become necessary. Several approximations have been developed to take into account the HIs. The more accurate descriptions are those that involve the calculation of the diffusion tensor,[30] mainly based on the Rotne–Prager–Yamakawa approach.[31–33] However, the calculation of the diffusion tensor at every time step is computationally very expensive. An alternative method is based on the mean field equations proposed by Tokuyama,[34–36] which introduce an effective diffusion coefficient accounting for the HIs. This effective diffusion coefficient is computed as a function of the volume fraction and the dilute solution diffusion coefficient. However, the Tokuyama method is limited since it was deduced for systems of equal sized hard-core spheres. In previous studies that use Tokuyama equations, macromolecules are considered as hard spheres and implicit solvent via Brownian Dynamics (BD) is used.[18,20–23]

In general, macromolecules have a flexible structure. As a result, when two macromolecules approach each other, their branches can become entangled and non-specific attractive and/or repulsive interactions between their chains are expected. This fact can play an important role in macromolecular diffusion in crowded media. Therefore, models going beyond the hard-core spheres, in which the conformational dynamic behaviour of macromolecules is lost, are necessary.[37] Softened interaction potentials have been previously proposed in the study of colloidal systems such as polymeric micelles[38,39] and star-polymers.[40–44] These systems consist of a dense core surrounded by a rather sparse corona. For pluronic polymeric micelles, studies based on implicit solvent and coarse grained computations have shown that both attractive and repulsive forces are present. However, the resulting potential of mean force between two particles is found to be purely repulsive.[39] Similar behaviour is found in the effective interparticle potentials used in star-polymer models. Star polymers can be modelled as surfaces coated with grafted polymers interacting by excluded volume effects. The particles are large enough to employ the Derjaguin approximation.[45–47] The proposed potential is the result of the soft repulsion

between the polymeric branches and a hard-core repulsion between the rigid cores.

In this paper, diffusion of the streptavidin protein among dextran molecules will be analysed. Dextran has a branched, highly hydrated, polymeric structure. Dextran molecules are significantly smaller than micelles and star-polymers, so that some approximations previously used in these systems, such as the Derjaguin approximation, are no longer valid. On the other hand, their very sparse structure and chemical composition, including polar alcoholic groups, can compensate in part the steric repulsions and favour macromolecular entanglement, so that a singular treatment becomes necessary.

In Section 2, a new coarse-grained approach is proposed: the Chain Entanglement Softened Potential model (CESP). The CESP model is based on an inter-particle interaction potential depicted in Fig. 1. It includes both hard-core and soft interaction regions at proper characteristic lengths. In the soft region, where the potential is shoulder-shaped, particle overlapping is possible, so that macromolecular branches are allowed to become entangled. A single parameter controls the entanglement energy. This potential is introduced in the BD scheme in which the HIs are implemented using the mean-field Tokuyama equations. D^{long} and α are obtained from the simulated temporal evolution of D by using a new empirical expression. The new equation, unlike the mostly used power law (eqn (1)), is able to describe the evolution of D for the three temporal regimes.

In Section 3, the CESP model is used to study streptavidin diffusion in crowded media among dextrans of different sizes and concentrations. Firstly, the shoulder shaped potential is
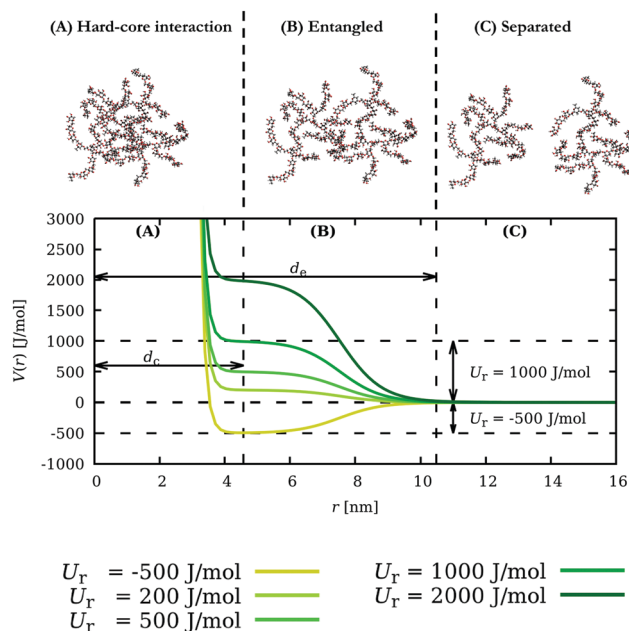

Fig. 1 Outline of the Chain Entanglement Softened Potential (CESP) model and the potential energy $V(r)$ vs. interparticle distance r (eqn (2)). Three interaction regions can be observed: no interaction (C), soft interaction where entanglement is allowed (B), and hard-core interaction (A). The chosen parameters are d_c = 4.58 nm and d_e = 10.48 nm, for U_r ranging from −500 to 2000 J mol^{-1}.

parametrized by fitting the computed D^{long} to the experimental values[11] for D50 (indicating its molecular weight in kg mol^{-1}) dextran obstacles. The BD simulations performed with the CESP model show a quantitative agreement between the calculated and experimental long time diffusion coefficient (D^{long}). This agreement is much better than the one obtained when the hard sphere model is used.[24] Once parametrized, we show that the CESP model is able to quantitatively predict D^{long} and the anomalous exponent (α) for the rest of the dextran sizes (D10, D400 and D700) in a wide range of macromolecular concentrations. Analysis of the radial distribution functions shows a significant increase in the population of entangled macromolecules with macromolecular concentration. This result highlights the importance of including the conformational dynamics of the macromolecules in the diffusion processes in crowded media.

2 Methodology

2.1 Chain Entanglement Softened Potential (CESP) model

In BD simulations of protein diffusion in crowded media, it is usual to reduce the resolution of the macromolecules to effective hard spheres.[18,20–23] This approximation significantly reduces the computational cost allowing larger time scales to be reached. However, the loss of the macromolecular structural properties could become unrealistic in many cases.[37] As the volume fraction increases, the macromolecules are expected to start to become entangled, a fact that cannot be properly described within the hard-core sphere model.[24]

In the present work, macromolecular entanglement is introduced *via* a coarse-grained model, the Chain Entanglement Softened Potential model (CESP). The model relies on an empiric interparticle interaction potential that resembles the continuous shouldered well potential used in computational models of coarse-grained fluids.[48,49] In these studies, density anomalies have been reported when a softened inter-particle potential is applied. The obtained radial distribution functions showed that an inner coordination shell is formed when the fluid density increases. Moreover, the population of the inner coordination shell was found to be proportional to the fluid density. Here, a redefinition of this idea is proposed. The coarse grained particles are no longer fluid molecules but macromolecules in solution. As the density of macromolecules in solution increases, they start to become entangled. Similar potentials have been used in the description of protein–surface interactions.[50]

The interaction potential corresponding to the CESP model is shown in Fig. 1. Two clear interaction regions can be observed that define two characteristic distances: the entanglement distance (d_{e}) and the core distance (d_{c}). If the macromolecules are at a distance r larger than d_{e}, they are considered to be separated, and they interact weakly (Fig. 1C). However, when the macromolecules are at $d_{\text{c}} \leq r \leq d_{\text{e}}$, they are considered to become entangled (Fig. 1B). In the latter situation, steric repulsions start to arise between the macromolecular chains. Such repulsions

can be compensated in part by the presence of van der Waals and hydrogen bonding attraction forces. As a result, a shoulder is developed in the proposed potential (Fig. 1). If the macromolecules get closer ($r < d_{\text{c}}$), the steric repulsion between the macromolecular skeletons dramatically increases and the hard-core region emerges (Fig. 1A). The resulting picture resembles models previously proposed for polymer micelles and star-polymers, composed of a dense core and a sparse corona. Unlike these systems, dextran molecules are polysaccharides with a complex branched structure where no structurally differentiated parts can be identified.[51,52] Therefore (d_{e}) and (d_{c}) should be regarded as the characteristic lengths of the different interaction regimes rather than chemically defined parts of the macromolecules. The proposed potential $V(r)$ acting between two particles at a distance r reads

$$V(r) = \varepsilon_0 \left(\frac{d_{\text{c}}}{r}\right)^n - \frac{U_{\text{r}}}{2}\tanh\left(\frac{d_{\text{c}}/d_0}{d_{\text{e}} - d_{\text{c}}}\left(r - \frac{d_{\text{e}} + d_{\text{c}}}{2}\right)\right) + \frac{U_{\text{r}}}{2} \quad (2)$$

where U_{r} is a parameter that quantifies the entanglement energetic cost. The sign of U_{r} indicates which forces, repulsive (positive sign) or attractive (negative sign), dominate in the soft interaction region. The first term in eqn (2) accounts for the hard-core repulsion region. Subsequently,[48–50] the exponent n is set to $n = 24$. $\varepsilon_0 = 1$ J mol^{-1} and $d_0 = 1$ nm are just to set the units. $V(r)$ is depicted in Fig. 1 for U_{r} values ranging from -500 to 2000 J mol^{-1}. Although the real situation is most probably more complex, eqn (2) can be regarded as a minimal model, compatible with the physicochemical properties of macromolecules considered, with only one parameter to be fitted.

The proposed interaction potential will be used to analyze the diffusion of streptavidin among different sized dextrans acting as inert obstacles,[11] tracked down using FCS. Both the tracer protein and the dextran obstacles are considered to be able to become entangled. The differences between protein/dextran and dextran/dextran (protein/protein interactions are not relevant due to the large excess of dextran) are then included in an effective way by the specific values of d_{c} and d_{e} for the protein and dextran, while the U_{r} value should be understood as an average interaction parameter. In order to estimate d_{e} and d_{c}, the interaction regions in a macromolecule are considered to be concentric spheres with two characteristic radii: the entanglement radius (R_{e}) and the hard-core radius (R_{c}), so that for two interacting particles i and j, $d_{\text{e},ij} = R_{\text{e},i} + R_{\text{e},j}$ and $d_{\text{c},ij} = R_{\text{c},i} + R_{\text{c},j}$.

In this work, R_{e} is associated with the hydrodynamic radius of the macromolecule in dilute solution. For dextran molecules, it has been computed by fitting experimental hydrodynamic radii *vs.* molecular weight obtained by quasi-elastic light scattering to an empiric power law function.[24,53] The dilute solution diffusion coefficient (D_0) for dextrans is calculated by incorporating the resulting R_{e} values into the Stokes–Einstein equation. On the other hand, the HYDROPRO software (Version 10)[54] has been used to compute D_0 corresponding to streptavidin. HYDROPRO considers each atom in the protein surface as a hydrodynamic frictional sphere. Protein atomic coordinates corresponding to three different crystallographic structures[55–57] were used in the calculations. D_0 was taken as the average of the

values resulting from the three structures. R_e of streptavidin is then obtained *via* the Stokes–Einstein equation.

Here, we propose to estimate R_c as the maximum approach distance between macromolecules, and it has been estimated using the specific volume ν, considering them as compact spheres[24]

$$R_c = \sqrt[3]{\frac{3\nu}{4\pi N_A} M_w} \qquad (3)$$

where N_A is the Avogadro number and M_w is the molecular weight. We have used $\nu = 0.625$ cm³ g⁻¹ for dextrans[58] and $\nu = 0.71$ cm³ g⁻¹ for streptavidin.[59]

M_w, D_0, R_e and R_c for streptavidin and different sized dextrans are reported in Table 1. Interestingly, the ratio R_e/R_c increases with dextran size, which could be caused by an increase in dextran branching with molecular weight. Note also that D10 and streptavidin have similar D_0 and R_e values. However, D10 has a higher ratio R_e/R_c than streptavidin, in accordance with the very different chemical structures of both macromolecules, much more compact in the last case.

2.2 Brownian Dynamics algorithm

Brownian Dynamics (BD) simulations make use of the Langevin equation in the over-damped limit,[60,61] so that the particles are considered to be under the effect of a stochastic force that accounts for the collisions with the solvent. The stochastic force is mathematically represented by Gaussian random noise,[19,24,25,62] and the equation of motion for the N particles reads

$$\mathbf{x}(t + \Delta t) = \mathbf{x}(t) - \frac{\Delta t}{RT} \mathbf{D} \nabla V(r) + \sqrt{2\mathbf{D}\Delta t}\xi(t) \qquad (4)$$

where ξ is a vector of $3N$ Gaussian random numbers with zero mean and unit variance, T is the system temperature, $\mathbf{x}$ is a vector with the $3N$ Cartesian coordinates, V is the potential of the CESP model (eqn (2)) and $\mathbf{D}$ is the diffusion tensor.

Since the solvent is simulated implicitly, the HIs must be included in eqn (4). HIs are solvent mediated correlations of the particle motions and they have been found to play an important role in macromolecular diffusion in crowded media.[24,28,63] HIs can be included in eqn (4) by calculating $\mathbf{D}$ following the Rotne–Prager–Yamakawa (RPY) method at each time step,[30] which is computationally very demanding.

Tokuyama equations[34–36] offer an alternative to the RPY method based on a mean field approximation. In this approach, an effective diffusion coefficient that accounts for the particle

mobility reduction due to the HIs is analytically calculated. The expression for the diffusion coefficient is deduced starting from first principles, *i.e.*, the Fokker–Planck equation for the single-particle distribution function coupled with the Navier–Stokes equations. This equation is analytically solved at short times considering a stationary configuration of equal-sized Brownian particles. In doing so, an approximation of the effective diffusion coefficient at short times (D^{short}) is obtained. For a particle diffusing in a system with a volume fraction ϕ, D^{short} reads

$$D^{short} = \frac{D_0}{1 + H(\phi)} \qquad (5)$$

where $H(\phi)$ describes the HI contributions in the short-time regime

$$H(\phi) = \frac{2b^2}{1-b} - \frac{c}{1+2c} - \frac{bc(2+c)}{(1+c)(1-b+c)} \qquad (6)$$

with $b = \sqrt{\frac{9}{8}\phi}$ and $c = \frac{11}{16}\phi$.

In this approach, $\mathbf{D}$ in eqn (4) is replaced by the effective D^{short}, which is a function of ϕ and D_0. However, the choice of ϕ presents two main difficulties. On the one hand, for a number of obstacles N_O, one could naively calculate ϕ as

$$\phi = \frac{4\pi N_O R_O^3}{3V} \qquad (7)$$

by identifying the obstacle radius R_O as the hydrodynamic radius. However, this leads to unphysical ϕ values larger than unity for the higher dextran concentrations, suggesting the existence of interpenetration between dextran molecules. On the other hand, the soft nature of the particles and the possibility of entanglement cause ϕ to be undetermined. A possible solution to this problem could be to interpret ϕ as an effective volume fraction ϕ^{eff}, depending on the macromolecular concentration and molecular size *via* some parameters to be determined. Then, the problem would be to find the value for these parameters that better predicts the experimental D^{long} using BD dynamics and Tokuyama equations. However, finding the best-fitted value of these parameters would require a repeated iteration of the BD dynamics, which is computationally very expensive. Instead, a heuristic approach has been chosen, consisting of estimating ϕ^{eff} using an analytic expression for D^{long} deduced by Tokuyama for equal sized hard-core spheres in the absence of the interaction force

$$D^{long} = \frac{D^{short}}{1 + \kappa \frac{D^{short}\phi^{eff}}{D_0 \phi_c}\left(1 - \frac{\phi^{eff}}{\phi_c}\right)^{-2}} \qquad (8)$$

where κ and ϕ_c are parameters set to $\kappa = 2.0$ and $\phi_c = 1.09$.[35] By equating eqn (8) to the experimental D^{long}, we can obtain a first approximation to ϕ^{eff}. The resulting ϕ^{eff} values have been found to depend not only on dextran concentration but on the dextran size. For the same dextran mass concentration, larger dextran sizes lead to larger ϕ^{eff} values, so that dextrans seem to effectively occupy more volume as the molecular size increases. In fact, the computed ϕ^{eff} values have been found to follow the

Table 1 Molecular weight (M_w) in kg mol⁻¹ (kDa),[11] entanglement radius (R_e) and hard core radius (R_c) for streptavidin and different sized dextrans (D10, D50, D400 and D700)

	M_w [kDa]	R_e [nm]	R_c [nm]	R_e/R_c	D_0 [nm² ns⁻¹]
Streptavidin	52.8	3.04	2.45	1.24	0.085
D10	11.6	2.77	1.43	1.94	0.089
D50	48.6	5.24	2.29	2.29	0.047
D400	409.8	13.52	4.67	2.90	0.018
D700	667.8	16.81	5.51	3.14	0.015

simple empiric equation (which resembles the Langmuir–Freundlich isotherm[64])

$$\phi^{\text{eff}} = \frac{z^m}{1 + z^m} \qquad (9)$$

with $z = k\nu c\dfrac{R_e}{R_c}$, where c is the dextran concentration, ν represents the specific value, and R_e/R_c is the characteristic radii ratio. k and m are fitted parameters whose values are $k = 3.7 \pm 0.1$ and $m = 1.43 \pm 0.08$. The best-fitted ϕ^{eff} as a function of z for D10 (squares), D50 (circles), D400 (upward-pointing triangles) and D700 (downward-pointing triangles) is shown in Fig. 2. The dashed line represents the best-fitted curve eqn (9). This procedure is actually expected to provide an approximation to ϕ^{eff}, whose suitability can only be justified a posteriori, if the BD simulations coupled to the CESP model properly reproduce the experimental D^{long} values. This is the case, as will be shown in the next section.

BD simulations are performed in a cubic box under periodic boundary conditions in all directions. The time step is set to 0.1 ns and the simulation time ranges from 10 to 150 μs depending on the time necessary to reach the stationary value of the diffusion coefficient at long times (D^{long}). The system temperature is 298.15 K. D is computed via eqn (5) and (6) using the ϕ^{eff} predicted by eqn (9) for each type of particle (dextran and protein). All the studied systems have 100 obstacle dextran particles and only one to five tracer protein particles to ensure that the volume occupied by the tracer is less than 1% of the volume occupied by the obstacles. The length of the simulation box is properly tuned to match the experimental dextran concentration and ranges from 17.56 nm (for the smaller dextran) to 164.32 nm (for the larger dextran). $\langle r^2 \rangle$ is computed after a thermalisation time of 0.5 μs in order to avoid possible initial unrealistic motion due to the random starting configuration. For each studied system, $\langle r^2 \rangle$ is obtained by averaging from 1000 (higher macromolecular concentration) to 15 000
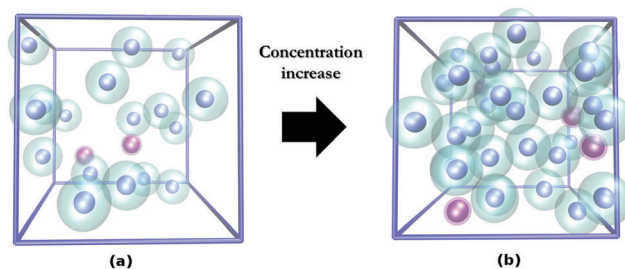

Fig. 3 Illustration of the simulation box of two BD simulations of the streptavidin protein (purple) diffusing among D50 dextran obstacles (blue). Dextran concentrations are (a) 25 g L^{-1} and (b) 100 g L^{-1}. Each macromolecule is represented as two concentric spheres representing the soft (transparent) and hard-core (opaque) interaction regions. As concentration increases, steric compression promotes macromolecules to become entangled.

(lower macromolecular concentration) BD realizations in order to ensure appropriate statistics.

In Fig. 3, two BD simulation boxes are depicted to illustrate the CESP model. Tracer streptavidin protein molecules (purple) diffuse among D50 dextran obstacles (blue). Each macromolecule is represented as two concentric spheres with radii of R_c and R_e, which characterize the soft (transparent) and the hard-core (opaque) interaction regions, respectively. The macromolecular concentrations of the solutions are 25 g L^{-1} (Fig. 3a) and 100 g L^{-1} (Fig. 3b). In the less concentrated solution, it can be observed that the soft interaction is enough to avoid macromolecule overlapping. In the more concentrated solutions, steric compression allows the macromolecules to overcome the initial steric repulsion and become entangled.

2.3 Determination of D^{long}, D^{short} and α from BD simulations: a new empirical equation for the time evolution of D

As mentioned before, the diffusion coefficient D of a particle is known to have a transition between two steady states: D^{short} and D^{long}. $\log(D)$ vs. $\log(t)$ profiles computed from BD simulations can be very well fitted to the function

$$\log(D) = B \tanh(\beta(\log(t) - \gamma)) + A \qquad (10)$$

where A, B, β and γ are parameters related to the relevant dynamic properties of the system (the mathematical details are given in the appendix): $A = \log(D^{\text{short}}D^{\text{long}})/2$, $B = \log(D^{\text{long}}/D^{\text{short}})/2$, $\beta = (\alpha - 1)/B$ and γ is the inflection time between D^{short} and D^{long}. This new empirical equation is able to describe the diffusion coefficient in the three temporal regimes. This fact represents an advantage compared to the scaling law (eqn (1)), which is unable to describe the asymptotic behaviour at short and long times. Moreover, it provides a simple method to compute D^{short}, D^{long} and α from the trajectory of a tracer particle.

In Fig. 4, the time evolution of the simulated diffusion coefficient of streptavidin in a D50 crowded solution, at a macromolecular concentration of 350 g L^{-1}, is shown (blue line). As expected, the diffusion coefficient undergoes a transition from D^{short} to D^{long} as macromolecular crowding slows down the motion of the tracer particle. It can be observed that eqn (10), depicted in red, fits very well to the computed values
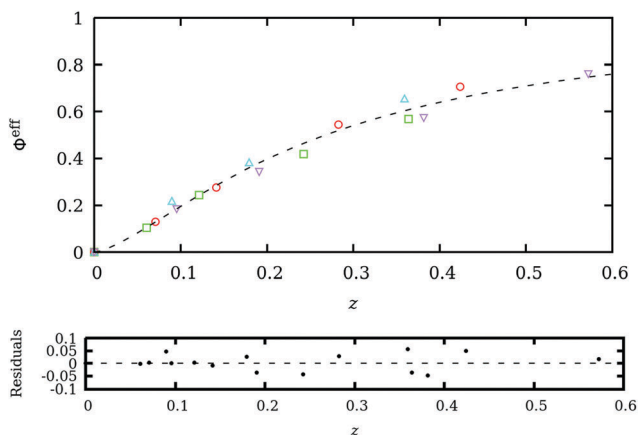

Fig. 2 Effective volume fraction (ϕ^{eff}) as a function of $z = k\nu cR_e/R_c$ estimated using eqn (5)–(8) for D10 (squares), D50 (circles), D400 (upward-pointing triangles) and D700 (downward-pointing triangles) dextran molecules. The dashed line represents the fitted eqn (9). The residuals of the fitting procedure are also depicted and show a very good fitting.
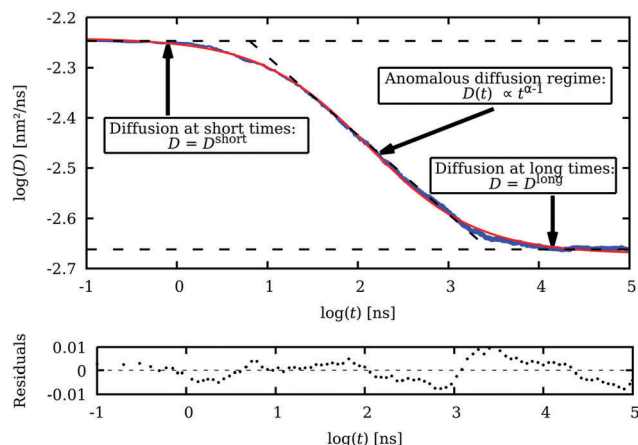
Fig. 4 Time evolution of the diffusion coefficient (D) in logarithmic scale. The blue line results from averaging 3000 BD simulations of 100 µs long in a cubic simulation box with a side of 28.466 nm. The simulation box contains 5 tracer particles (streptavidin) diffusing among 100 obstacle particles (D50 dextran). The macromolecular concentration is 350 g L^{-1}. Macromolecular crowding slows down the tracer diffusion coefficient from D^{short} to D^{long}. The red line is the result of fitting eqn (10) to the simulated results, which allows D^{short}, D^{long} and α to be easily determined. The resulting fitting is very good, as shown by the residuals.

over all the temporal regimes, allowing rigorous and easy determination of D^{short}, D^{long} and α.

3 Results and discussion

3.1 Model parametrization

The proposed shoulder-shaped potential used in the CESP model includes only one parameter to be fitted, *i.e.*, the entanglement energetic cost U_r (eqn (2)). In order to properly set U_r, BD simulations have been performed with U_r values ranging from 200 to 2000 J mol^{-1}. In these simulations, D^{long} of streptavidin (normalised to D_0) has been computed at seven macromolecular concentrations of D50 dextran: 50, 100, 200, 210, 260, 300 and 360 g L^{-1}. D^{long} is obtained by fitting the computed $\log(D)$ *vs.* $\log(t)$ profile to eqn (10).

The obtained values are shown in Fig. 5 in green colour gradation. As can be observed, D^{long} decreases with U_r because of the increase in the impenetrability of the particles and the consequent reduction in the available space for the particle to diffuse. Experimental D^{long} values reported in ref. 11 are also depicted in the same figure (red markers). For U_r = 500 J mol^{-1}, very good quantitative agreement between simulations and experiments is obtained. In all the BD computations performed in the rest of this work, U_r has been set to this value, which indicates that, for dextran, the repulsion forces arising from chain entanglement slightly dominate over the attractive ones. D^{long} values obtained using the hard sphere model, reported in ref. 24, are also depicted in Fig. 5 (black markers). In this previous computational work, dextran macromolecules were modeled considering the dextran molecules as hard-core spheres with an effective radius, which was chosen as an intermediate value lying between the hydrodynamic radius[53] and the compact radius.
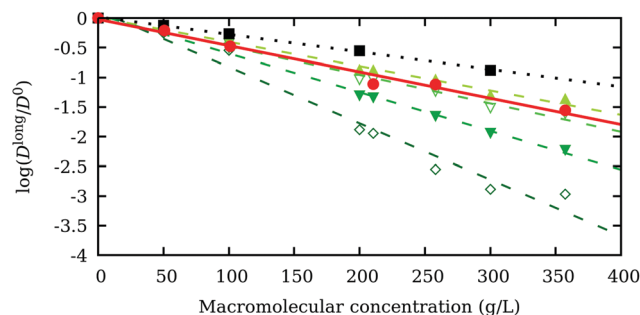

Fig. 5 Calculated D^{long}/D_0 as a function of the concentration corresponding to D50 dextran for U_r values: 200 J mol^{-1} (green triangles), 500 J mol^{-1} (empty green inverted triangles), 1000 J mol^{-1} (green inverted triangles) and 2000 J mol^{-1} (green diamonds). Black square markers represent the values obtained using the hard sphere model (taken from ref. 24) while red circle markers denote the experimental D^{long} value (taken from ref. 11). For U_r = 500 J mol^{-1}, good agreement between the CESP model and the experimental values is observed. Lines are only to guide the eye.

This approach accounted for the possibility of macromolecular overlapping in an average way, but represented a static description of the conformational state, being unable to properly mimic macromolecular entanglement. As is clearly observed in Fig. 5, D^{long} values obtained using the CESP model are much closer to the experimental ones than those coming from the hard-core sphere model.[24] For the rest of the dextran sizes (D10, D400 and D700), direct comparison of D^{long} predicted by the hard-sphere model[24] and the CESP model can be found in the ESI.† Again, it is found that the CESP model shows a significant improvement in the prediction of the obtained experimental D^{long} over that using the hard-core sphere model. Therefore, the inclusion of coupling of the dynamic and conformational properties seems to be necessary to accurately describe diffusion in crowded media.

3.2 Model prediction for D^{long} and the anomalous exponent α

In the previous section, we have found that for U_r = 500 J mol^{-1}, the CESP model is able to describe D^{long} of streptavidin for D50 dextran obstacles. Let us now see that this model can also predict, without any further parametrization, D^{long} and α of streptavidin for different dextran sizes (D10, D400 and D700) and concentrations (from 25 to 350 g L^{-1}). D^{long} and α are again obtained from the BD simulations by fitting eqn (10) to the computed D time evolution. The curve $\log(D)$ *versus* $\log(t)$ has been plotted in Fig. 6, together with the best fitted eqn (10), for four dextran sizes and concentrations: (Fig. 6a) D10 at 300 g L^{-1}, (Fig. 6b) D50 at 200 g L^{-1}, (Fig. 6c) D400 at 100 g L^{-1} and (Fig. 6d) D700 at 50 g L^{-1}. As can be observed, for all the systems presented, eqn (10) fits very well to the computational results.

The computed D^{long} *versus* macromolecular concentration plots are depicted in Fig. 7 for four dextran sizes: D10, D50, D400 and D700. The results are shown together with the experimental values reported in ref. 11 for the same systems. Dashed lines are only to guide the eye and they follow the simulation results. In general, the computed results are in very good quantitative
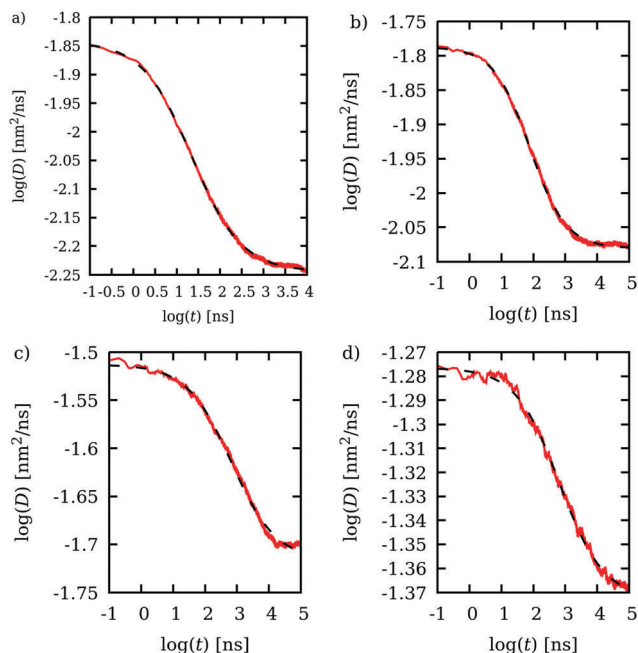
Fig. 6 Time evolution of the diffusion coefficient D versus time in the logarithmic scale obtained from BD computations (red line). The best-fitted eqn (10) is plotted as a black dashed line. Four cases are depicted: (a) D10 at 300 g L^{-1}, (b) D50 at 200 g L^{-1}, (c) D400 at 100 g L^{-1} and (d) D700 at 50 g L^{-1}. For all the systems presented, eqn (10) fits very well to the computed values, allowing easy determination of D^{long}, D^{short} and α from the temporal evolution of D.
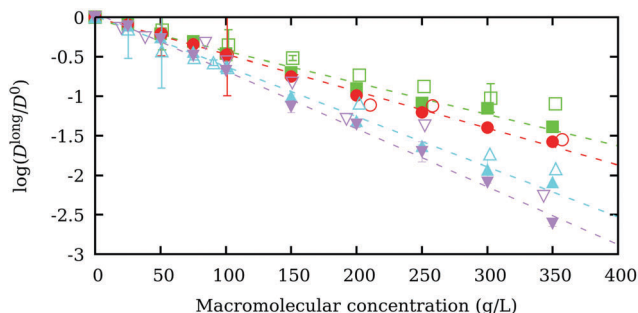

Fig. 7 $\log(D^{long}/D_0)$ for streptavidin versus macromolecular concentration. The filled markers represent BD simulations and the empty ones correspond to experimental values[11] for four dextran sizes: D10 (green squares), D50 (red circles), D400 (upward pointing blue triangles) and D700 (downward pointing purple triangles). Dashed lines are only to guide the eye and they follow the simulation results.

agreement with the experimental values. However, the values corresponding to D10 exhibit some systematic error, which could be explained taking into account that the potential energy barrier U_r has been set, regardless of the dextran size, to the best fit D^{long} for D50. This suggests that U_r should be smaller for D10 than for D50, D400 and D700. Since U_r accounts for the energy cost of the macromolecular entanglement, this fact could suggest a lower energy cost for D10 entanglement than the one for D50, D400 and D700. In agreement with the experiments reported in ref. 11, BD simulations predict a
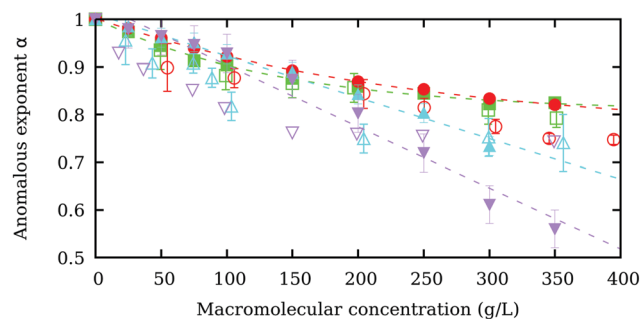

Fig. 8 Anomalous exponent α versus macromolecular concentration. Filled markers represent BD simulations and empty markers correspond to the experimental values[11] for four dextran sizes: D10 (green squares), D50 (red circles), D400 (upward pointing blue triangles) and D700 (downward pointing purple triangles). Dashed lines are only to guide the eye and they follow the simulation results.

decrease in D^{long} with macromolecular concentration, as a result of volume exclusion and hydrodynamic interactions. Alternative potentials based on a self-consistent field method appropriate for weak excluded interactions[45] have been implemented in BD simulations of star-polymer micelles.[42] As in dextran molecules, the computational results exhibit similar D^{long} decay with the particle density. However, at the quantitative level, the calculated values are much larger than the experimental ones. The discrepancies could be explained by the fact that HIs were not included in the computations of ref. 42. In the present work, HIs are included in the BD scheme, and it has been found that they are crucial for the quantitative reproduction of the experimental D^{long} values.[24,28]

The anomalous exponent α provides information about the transition between the short and the long-time regimes. α values close to unity indicate a smooth transition. As α deviates from unity, the transition becomes more pronounced. The simulated α versus macromolecular concentration plots are depicted in Fig. 8 together with the experimental values obtained using FCS.[11] As can be observed, the CESP model qualitatively captures the experimentally observed behaviour. α decays with macromolecular concentration as a result of an increase in particle collisions. Moreover, α also decreases with dextran size, which could be a consequence of the increase of ϕ^{eff}, so that the volume fraction of the system seems to play a crucial role in α and therefore in the transition between the D^{short} and D^{long} steady states. However, quantitative discrepancies are observed, which could be explained not only by the limitations of the model, but by the way in which α has been calculated. The experimental α values have been obtained by fitting the auto-correlation function of the tracer protein fluorescence using the power law described in eqn (1) while here, α has been determined as the slope in the inflection point of the curve $\log(D)$ versus $\log(t)$ using eqn (10).

3.3 Radial distribution functions

The simulated radial distribution functions (rdfs) for the cases analyzed in the previous section (D10 (a), D50 (b), D400 (c) and
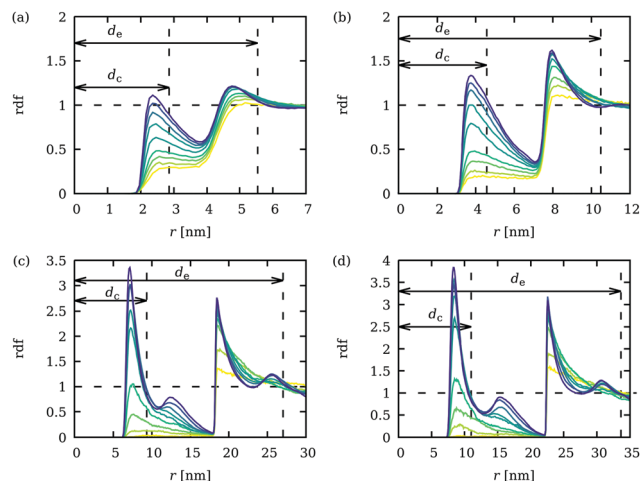
Fig. 9 Simulated radial distribution functions (rdfs) *versus* inter-particle distance using the CESP model for four dextran sizes: (a) D10, (b) D50, (c) D400 and (d) D700. For each size, the rdf has been computed at eight concentrations ranging from 25 to 350 g L^{-1}, depicted in color gradation from yellow (25 g L^{-1}) to purple (350 g L^{-1}).

Fig. 10 Number of entangled macromolecules around a reference particle (N_E) *versus* macromolecular concentration for four dextran sizes: D10 (green squares), D50 (red circles), D400 (blue upward pointing triangles) and D700 (purple downward pointing triangles). The number of entangled particles increases with macromolecular concentration as a result of steric compression.

D700 (d)) are shown in Fig. 9 for macromolecular concentrations ranging from 25 g L^{-1} to 350 g L^{-1}. The rdfs have been computed from the BD computations by calculating the center-to-center distance of every dextran obstacle to the rest of the dextran particles. Averages have been taken over 10 000 configurations every 2 ns of simulation. As can be observed, the CESP model predicts an increase in the number of entangled macromolecules with macromolecular concentration, as revealed by the gradual increase in the rdf at $r = d_c$. This fact is observed for all the studied dextran sizes. Two sharp peaks at distances slightly lower than d_c and d_e define two coordination shells of entangled macromolecules, corresponding to highly entangled macromolecules (close to d_c) and weakly entangled macromolecules (close to d_e), respectively. It is also interesting to note the presence of two secondary peaks for D400 and D700 at macromolecular concentrations of 250, 300 and 350 g L^{-1}, corresponding to additional coordination shells and suggesting other possible degrees of entanglement for the biggest dextran molecules. Note that the two main peaks in the entanglement region are developed even in the absence of explicit attractive terms in the proposed potential. Entanglement is thus the way the system can minimize the excluded volume effects, even paying a small energetic cost. It is reasonable to conclude that the presence of attractive forces would enhance this effect, although it is not necessary to produce it.

The first rdf peak has been integrated in order to determine the number of entangled particles (N_E) close to the core of the reference particle. N_E is shown in Fig. 10 for the four analyzed dextran sizes. It can be observed that N_E increases exponentially with macromolecular concentration, so that entanglement is promoted by steric compression. This effect is enhanced by the increase in dextran size, as a consequence of the increase in ϕ^{eff}, so that bigger dextran obstacles are able to effectively occupy higher volumes than the smaller ones for the same macromolecular concentration.
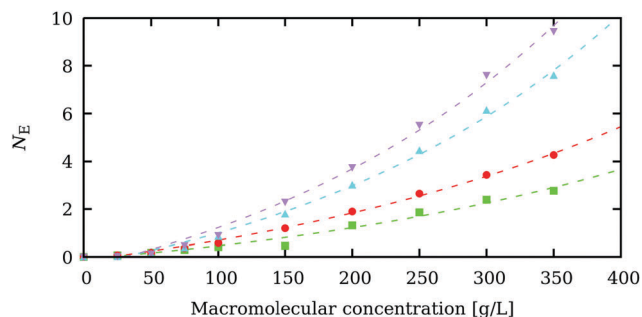
4 Conclusions

A new coarse grained approach to macromolecular diffusion in crowded media that goes beyond the hard sphere model is presented: the Chain Entanglement Softened Potential (CESP) model. Two interaction regions between macromolecules are considered: a soft interaction region caused by chain entanglement, a combination of repulsive and attractive effects, and a hard-core interaction region caused by the dramatic increase in steric repulsions between macromolecular skeletons when they approach below a certain distance. The resulting picture is quantified by means of a shoulder-shaped potential that allows macromolecular conformational flexibility and chain entanglement to be taken into account. The model can be parametrized by a unique parameter (U_r) that is associated with the entanglement energetic cost. For streptavidin protein diffusion in D50 crowded solution, U_r has been quantified by fitting the long time diffusion coefficient (D^{long}) resulting from BD simulations to the experimental values obtained using Fluorescence Correlation Spectroscopy (FCS).[11]

In order to determine D^{long} and the anomalous exponent α from BD simulations, a new empiric expression for the temporal evolution of the diffusion coefficient, valid for all the temporal regimes, has been proposed. The new function goes beyond the standard power law and depends on the four relevant dynamic properties: D^{short}, D^{long}, α and the inflection time γ. This function fits very well to the BD computations for all the obstacle concentrations and sizes under study, and it could be useful in forthcoming computational studies focused on anomalous diffusion.

The computed D^{long} using the CESP model clearly improves on the results obtained in previous studies based on hard-core spheres.[24] Without any further parametrization of $V(r)$, the CESP model is able to provide quantitative predictions of D^{long} and α of streptavidin in media crowded by D10, D400 and D700 at several dextran concentrations (25 to 350 g L^{-1}). Radial distribution functions exhibit a significant increase in the population of entangled macromolecules with the macromolecular concentration and dextran size. Our results indicate that an adequate

description of macromolecular entanglement is essential to understand macromolecular diffusion in crowded media.

Conflicts of interest

The authors declare no conflicts of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix

The new empirical equation proposed here for the time evolution of D (eqn (10)) contains four parameters that are related to the relevant dynamic properties of the system, i.e., D^{short}, D^{long}, α and the inflection time γ. At short times ($\log(t) \to -\infty$), eqn (10) must fulfill $\log(D) \to \log(D^{\text{short}})$, while at long times ($\log(t) \to +\infty$), we have $\log(D) \to \log(D^{\text{long}})$. Taking limits in eqn (10)

$$\lim_{t \to -\infty} \log(D) = -B + A = \log\left(D^{\text{short}}\right) \qquad (11)$$

$$\lim_{t \to +\infty} \log(D) = B + A = \log\left(D^{\text{long}}\right) \qquad (12)$$

from which it follows that $A = \frac{1}{2}\log\left(D^{\text{short}} D^{\text{long}}\right)$ and $B = \frac{1}{2}\log\left(D^{\text{long}}/D^{\text{short}}\right)$. By construction, γ corresponds to the inflection point of the time evolution curve. Some elementary algebra shows that

$$\left(\frac{\text{d}^2 \log(D)}{\text{d} \log(t)^2}\right)_{\log(t)=\gamma} = 0 \qquad (13)$$

Finally, β can be related to the anomalous diffusion exponent α, used when the transition regime is analyzed in terms of the standard power law (eqn (1))

$$D = \Gamma_\alpha t^{\alpha-1} \Leftrightarrow \log(D) = (\alpha - 1)\log(t) + \log(\Gamma_\alpha) \qquad (14)$$

By taking the first derivative in eqn (10) and comparing to eqn (14), one obtains

$$\left(\frac{\text{d} \log(D)}{\text{d} \log(t)}\right)_{\log(t)=\gamma} = \beta \quad B = \alpha - 1 \qquad (15)$$

which leads to $\beta = (\alpha - 1)/B$.

Acknowledgements

References

1 A. P. Minton, *Biopolymers*, 1981, **20**, 2093–2120.
2 S. B. Zimmerman and S. O. Trach, *J. Mol. Biol.*, 1991, **222**, 599–620.
3 A. Minton, *Biophys. J.*, 1992, **63**, 1090–1100.
4 R. J. Ellis, *Trends Biochem. Sci.*, 2001, **26**, 597–604.
5 H.-X. Zhou, G. Rivas and A. P. Minton, *Annu. Rev. Biophys.*, 2008, **37**, 375–397.
6 C. Balcells, I. Pastor, E. Vilaseca, S. Madurga, M. Cascante and F. Mas, *J. Phys. Chem. B*, 2014, **118**, 4062–4068.
7 I. Pastor, L. Pitulice, C. Balcells, E. Vilaseca, S. Madurga, A. Isvoran, M. Cascante and F. Mas, *Biophys. Chem.*, 2014, **185**, 8–13.
8 C. Balcells, I. Pastor, L. Pitulice, M. Via, S. Madurga, E. Vilaseca, A. Isvoran, M. Cascante and F. Mas, *New Frontiers in Chemistry*, 2015, **24**, 3–16.
9 G. Rivas and A. P. Minton, *Trends Biochem. Sci.*, 2016, **41**, 970–981.
10 M. C. Konopka, I. A. Shkel, S. Cayley, M. T. Record and J. C. Weisshaar, *J. Bacteriol.*, 2006, **188**, 6115–6123.
11 D. S. Banks and C. Fradin, *Biophys. J.*, 2005, **89**, 2960–2971.
12 I. Pastor, E. Vilaseca, S. Madurga, M. Cascante and F. Mas, *J. Phys. Chem. B*, 2010, **114**, 4028–4034.
13 I. Pastor, E. Vilaseca, S. Madurga, M. Cascante and F. Mas, *J. Phys. Chem. B*, 2010, **114**, 12182.
14 J. A. Dix and A. Verkman, *Annu. Rev. Biophys.*, 2008, **37**, 247–263.
15 M. Feig, I. Yu, P.-h. Wang, G. Nawrocki and Y. Sugita, *J. Phys. Chem. B*, 2017, 1–17.
16 E. Vilaseca, A. Isvoran, S. Madurga, I. Pastor, J. L. Garcés and F. Mas, *Phys. Chem. Chem. Phys.*, 2011, **13**, 7396–7407.
17 E. Vilaseca, I. Pastor, A. Isvoran, S. Madurga, J. L. Garcés and F. Mas, *Theor. Chem. Acc.*, 2011, **128**, 795–805.
18 J. Sun and H. Weinstein, *J. Chem. Phys.*, 2007, **127**, 155105.
19 P. Mereghetti and R. C. Wade, *J. Phys. Chem. B*, 2012, **116**, 8523–8533.
20 S. Hasnain, C. L. McClendon, M. T. Hsu, M. P. Jacobson and P. Bandyopadhyay, *PLoS One*, 2014, **9**, 1–12.
21 S. Kondrat, O. Zimmermann, W. Wiechert and E. V. Lieres, *Phys. Biol.*, 2015, **12**, 046003.
22 T. Sentjabrskaja, E. Zaccarelli, C. De Michele, F. Sciortino, P. Tartaglia, T. Voigtmann, S. U. Egelhaaf and M. Laurati, *Nat. Commun.*, 2016, **7**, 11133.
23 I. Yu, T. Mori, T. Ando, R. Harada, J. Jung, Y. Sugita and M. Feig, *eLife*, 2016, **5**, 1–22.
24 P. M. Blanco, M. Via, J. L. Garcés, S. Madurga and F. Mas, *Entropy*, 2017, **19**, 105.
25 S. Smith and R. Grima, *J. Chem. Phys.*, 2017, **146**, 024105.
26 S. Bucciarelli, J. S. Myung, B. Farago, S. Das, G. A. Vliegenthart, O. Holderer, R. G. Winkler, P. Schurtenberger, G. Gompper and A. Stradner, *Sci. Adv.*, 2016, **2**, e1601432.
27 P. H. Wang, I. Yu, M. Feig and Y. Sugita, *Chem. Phys. Lett.*, 2017, **671**, 63–70.
28 T. Ando, E. Chow and J. Skolnick, *J. Chem. Phys.*, 2013, **139**, 121922.

29 M. Saxton, *Biophys. J.*, 1994, **66**, 394–401.

30 D. L. Ermak and J. A. McCammon, *J. Chem. Phys.*, 1978, **69**, 1352–1360.

31 J. Rotne and S. Prager, *J. Chem. Phys.*, 1969, **50**, 4831–4837.

32 H. Yamakawa, *J. Chem. Phys.*, 1970, **53**, 436–443.

33 J. Brady, *Annu. Rev. Fluid Mech.*, 1988, **20**, 111–157.

34 M. Tokuyama and I. Oppenheim, *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.*, 1994, **50**, 6–9.

35 M. Tokuyama, *Phys. A*, 2006, **364**, 23–62.

36 M. Tokuyama, T. Moriki and Y. Kimura, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2011, **83**, 1–8.

37 A. Irbäck and S. Mohanty, *Eur. Phys. J.-Spec. Top.*, 2017, **226**, 627–638.

38 D. Bedrov, C. Ayyagari and G. D. Smith, *J. Chem. Theory Comput.*, 2006, **2**, 598–606.

39 D. Bedrov, G. D. Smith and J. Yoon, *Langmuir*, 2007, **23**, 12032–12041.

40 C. N. Likos, H. Löwen, A. Poppe, L. Willner, J. Roovers, B. Cubitt and D. Richter, *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.*, 1998, **58**, 6299–6307.

41 C. N. Likos, H. Löwen, M. Watzlawek, B. Abbas, O. Jucknischke, J. Allgaier and D. Richter, *Phys. Rev. Lett.*, 1998, **80**, 4450–4453.

42 B. Loppinet, G. Fytas, D. Vlassopoulos, C. N. Likos, G. Meier and J. L. Guo, *Macromol. Chem. Phys.*, 2005, **206**, 163–172.

43 L. Liu, W. K. den Otter and W. J. Briels, *Soft Matter*, 2014, **10**, 7874–7886.

44 E. Locatelli, B. Capone and C. N. Likos, *J. Chem. Phys.*, 2016, **145**, 174901.

45 S. T. Milner, T. A. Witten and M. E. Cates, *Macromolecules*, 1988, **21**, 2610–2619.

46 J. Mewis, W. J. Frith, T. A. Strivens and W. B. Russel, *AIChE J.*, 1989, **35**, 415–422.

47 U. Genz, B. D'Aguanno, J. Mewis and R. Klein, *Langmuir*, 1994, **10**, 2206–2212.

48 P. Vilaseca and G. Franzese, *J. Chem. Phys.*, 2010, **133**, 1–11.

49 F. Leoni and G. Franzese, *J. Chem. Phys.*, 2014, **141**, 174501.

50 P. Vilaseca, K. A. Dawson and G. Franzese, *Soft Matter*, 2012, 6978–6985.

51 J. W. Sloan, B. H. Alexander, R. L. Lohmar, I. A. Wolff and C. E. Rist, *J. Am. Chem. Soc.*, 1954, **76**, 4429–4434.

52 J. Sabatie, L. Choplin, M. Moan, J. L. Doublier, F. Paul and P. Monsan, *Carbohydr. Polym.*, 1988, **9**, 87–101.

53 J. Armstrong, R. Wenby, H. Meiselman and T. Fisher, *Biophys. J.*, 2004, **87**, 4259–4270.

54 A. Ortega, D. Amorós and J. G. De La Torre, *Biophys. J.*, 2011, **101**, 892–898.

55 M. Fairhead, D. Krndija, E. D. Lowe and M. Howarth, *J. Mol. Biol.*, 2014, **426**, 199–214.

56 T. Terai, M. Kohno, G. Boncompain, S. Sugiyama, N. Saito, R. Fujikake, T. Ueno, T. Komatsu, K. Hanaoka, T. Okabe, Y. Urano, F. Perez and T. Nagano, *J. Am. Chem. Soc.*, 2015, **137**, 10464–10467.

57 L. Baugh, I. Le Trong, P. S. Stayton, R. E. Stenkamp and T. P. Lybrand, *Biochemistry*, 2016, **55**, 5201–5203.

58 D. W. Schaefer, *Polymer*, 1984, **25**, 387–394.

59 A. Pähler, W. A. Hendrickson, M. A. G. Kolks, C. E. Argaña and C. R. Cantor, *J. Biol. Chem.*, 1987, **262**, 13933–13937.

60 H. Mori, *Prog. Theor. Phys.*, 1965, **33**, 423–455.

61 D. T. Gillespie, *Am. J. Phys.*, 1996, **64**, 1246–1257.

62 J. K. Dhont, *An Introduction to Dynamics of Colloids*, Elsevier, Amsterdam, The Netherlands, 1996, ch. 2.

63 R. J. Phillips, J. F. Brady and G. Bossis, *Phys. Fluids*, 1988, **31**, 3462–3472.

64 J. L. Garcés, F. Mas, J. Cecília, E. Companys, J. Galceran, J. Salvador, J. Puy and J. Galceran, *Phys. Chem. Chem. Phys.*, 2002, **4**, 3764–3773.

This journal is © The Royal Society of Chemistry 2018