## PAPER

# Beads on a string: structure of bound aggregates of globular particles and long polymer chains

Anton Souslov,*[a] Jennifer E. Curtis[ab] and Paul M. Goldbart[a]

Macroscopic properties of suspensions, such as those composed of globular particles (*e.g.*, colloidal or macromolecular), can be tuned by controlling the equilibrium aggregation of the particles. We examine how aggregation – and, hence, macroscopic properties – can be controlled in a system composed of both globular particles and long, flexible polymer chains that reversibly bind to one another. We base this on a minimal statistical mechanical model of a single aggregate in which the polymer chain is treated either as ideal or self-avoiding, and, in addition, the globular particles are taken to interact with one another *via* excluded volume repulsion. Furthermore, each of the globular particles is taken to have one single site to which at most one polymer segment may bind. Within the context of this model, we examine the statistics of the equilibrium size of an aggregate and, thence, the structure of dilute and semidilute suspensions of these aggregates. We apply the model to biologically relevant aggregates, specifically those composed of macromolecular proteoglycan globules and long hyaluronan polymer chains. These aggregates are especially relevant to the materials properties of cartilage and the structure–function properties of perineuronal nets in brain tissue, as well as the pericellular coats of mammalian cells.

## I Introduction

For suspensions composed of long polymer chains and colloidal particles, a wealth of materials properties can be explored by tuning the polymer–colloid interactions. For example, in suspensions in which the polymers and the colloids do not bind to one another, flocculation of large aggregates can be controlled by tuning the polymer concentration, which determines the effective depletion or electrostatic forces between colloidal particles (as reviewed in, *e.g.*, ref. 1 and 2). In other systems, polymer–colloid interactions lead to the formation of polymer bridges between colloidal particles, which results in macroscopic gelation or flocculation (as reviewed in, *e.g.*, ref. 3). A similar bridging mechanism leads to gelation in polymer-micelle systems.[4–6] Such bridging phenomena have been modeled in terms of the adsorption of polymer chains onto the surfaces of suspended particles.[7,8] In this work, we instead consider a suspension in which polymer bridges form between globular suspended particles due to the presence on each such particle of a single specific site to which the polymer binds. For this system, aggregation is limited by the polymer chain length and, therefore, the aggregates are limited to the microscale and, thus, do not flocculate. We consider the case for which the polymer–colloid binding is reversible. For this

case, we describe how it may be possible to control experimentally the characteristic size of aggregates, for example by tuning the length of the polymer chains, the temperature, or the density of colloidal particles. The size of aggregates then determines the macroscopic materials properties of the suspension.

To develop a statistical mechanical approach to the structure of these aggregates, we model the colloidal particles as hard spheres, and the polymers as Gaussian chains, possibly with self-avoidance. By considering key physical features of polymer–colloid suspensions – the steric repulsion between the colloidal particles and the presence on each particle of exactly one binding site for polymer chains – we derive results for the statistics of the number $M$ of colloidal particles attached to each polymer chain and, consequently, of the linear size $R$ of the aggregate. Thus, for self-avoiding polymer chains we derive the following expressions for the mean and variance of $M$:

$$\langle M \rangle \sim (\beta(\mu + E_b))^{2/3}(R_g/2a)^{5/3} = (\beta(\mu + E_b))^{2/3}N(\ell_0/2a)^{5/3}, \quad (1)$$

$$\langle(\delta M)^2\rangle \sim (\beta(\mu + E_b))^{-1/3}(R_g/2a)^{5/3}, \quad (2)$$

in terms of $\beta = 1/k_B T$ (where $T$ is the temperature), the chemical potential $\mu$ of the colloidal particles, the polymer–colloid binding energy $E_b$, the characteristics of an isolated polymer chain (*i.e.*, the segment length $\ell_0$, the number of segments $N$, and the radius of gyration $R_g = N^{3/5}\ell_0$), and the radius $a$ of a colloidal particle. The relative width of the distribution [which is given by $\langle(\delta M)^2\rangle/\langle M\rangle^2 \sim (\beta(\mu + E_b))^{-5/3}$] decreases as the binding energy (and, thus, the size of the aggregate) increases. As a consequence,

[a] *School of Physics Georgia Institute of Technology, Atlanta, GA, 30332, USA.*
 *E-mail: anton.souslov@physics.gatech.edu*
[b] *Parker H. Petit Institute for Bioengineering and Bioscience,*
 *Georgia Institute of Technology, Atlanta, GA, 30332, USA*

for sufficiently large aggregates the size distribution is sharply peaked around the mean, and the aggregates may be addressed within one of the two essentially equivalent ensembles – canonical at fixed $M$, or grand canonical at fixed $\mu$. In suspension, such aggregates may themselves be modeled in terms of effective self-avoiding chains, each composed of $M$ segments of segment length $2a$, so that the typical linear size $R$ of an aggregate is given by $aM^{3/5}$.

The physical system that inspired this model is the suspension of biomolecular aggregates composed of long chains of the polysaccharide hyaluronan (HA, which behaves as a polymer chain[9]) and the types of globular macromolecules called proteoglycans (PG, *e.g.*, aggrecan, versican, brevican and neurocan – collectively known as lecticans[10]) that bind to HA. Although the "beads on a string" language has been used to describe the euchromatin structure of nucleosome beads along a string of DNA, we do not intend to describe that system in this work. These aggregates are found in numerous biological contexts, including the extracellular matrix (ECM)[11,12] and the cell-surface-associated polymer matrix called the pericellular coat.[13,14] For example, in cartilage, a high concentration of aggrecan (which is bound to hyaluronan strands and therefore aggregated) works in concert with a dense network of collagen fibrils to form a tough and resilient load-bearing material. The aggrecan attracts water into the tissue by osmosis, thereby exerting a swelling pressure on the collagen network and enabling the cartilage to resist large compressive loads (see *e.g.*, ref. 15). As another example, when bound to the surfaces of cells, PG aggregates co-regulate cell behavior, especially processes that involve changes in cell adhesion, such as migration,[16] proliferation,[17] and synaptogenesis.[18]

In this biological context, the parameters in eqn (2) take on specific values or ranges of values:

• For HA chains, we choose the segment length to be twice the persistence length, *i.e.*, $\ell_0 \sim 14$ nm,[9] and the contour length $N\ell_0$ to be up to 25 μm,[11] such that $N \lesssim 10^3$ segments.

• We choose $a \sim 100$ nm to correspond to the size of the PG molecules that attach to the HA chain. In real systems, the PG molecules are somewhat anisotropic bottlebrushes, having cylindrical dimensions $\sim 100$ nm $\times$ 100 nm $\times$ 350 nm when fully stretched[19,20] but, for simplicity, we do not include this anisotropy in the model. We expect that, as a consequence of thermal fluctuations, the typical shape of the PG molecules in solution is less anisotropic. In order to model globular bottle-brush polymers that are highly swollen as a result of electrostatic repulsion between their highly charged side-chains (see Fig. 1),[21,22] we include hard-core repulsion between the globules. In biological contexts, the Debye length is sufficiently short ($\sim 1$ nm) that the only relevant interaction is the strong, steric repulsion that suppresses PG–PG overlap.

• The binding site of each PG molecule occupies $\sim 10$ nm of HA contour length, *i.e.*, a length $\sim \ell_0$.[23]

• PG–HA aggregates in cartilage are composed of up to $\sim 100$ PG globules bound to a single HA chain.[23]

## II Model and results

We begin by giving a detailed account of the models we shall use to describe a composite system consisting of a single polymer chain and the globular particles that bind to it. Next, we construct the partition function for a single aggregate. We then simplify this partition function to obtain our main result: the distribution of aggregate sizes, by which we mean the distribution of $M$ as well as of $R$.

We emphasize the interplay between chain entropy and the binding statistics of globules, and also emphasize associated phenomena occurring on lengthscales that are much larger than the HA persistence length. At these large lengthscales, a chain model is sufficient to capture the physical properties of the polymer, and we consider both the models of Gaussian and
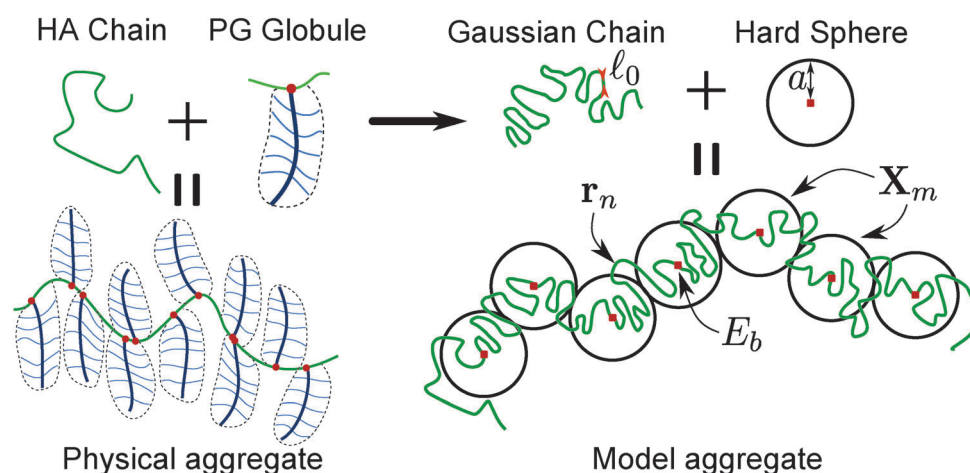


Fig. 1 A schematic diagram of the model presented in this work and its relation to molecular aggregates composed of hyaluronan (HA) and proteoglycans (PG). The variables $\mathbf{r}_n$ and $\mathbf{X}_m$ correspond to the positions of the chain segments and the globule centers, respectively. The model parameters include $E_b$ – the binding energy, $\ell_0$ – the chain segment length, and $a$ – the globule size. The PG globules, which are bottle-brush polymers that avoid overlap, are modeled as hard spheres and the polysaccharide HA chain is modeled as a Gaussian chain. These two components aggregate due to the affinity of the globules to bind to the chain.

self-avoiding chains. For a Gaussian chain, the polymer segments are modeled by Hookean springs (of natural length $\ell_0$), which connect universal joints at positions $\mathbf{r}_n$ (with $n = 1, \ldots, N'$). Self-avoiding chains, in addition, self-interact, *via* the potential $U_{n,n'} \equiv U(|\mathbf{r}_n - \mathbf{r}_{n'}|)$, which is relevant for physical polymers in a good solvent. Thus, the effective free energy of chain configuration $\{\mathbf{r}_n\}$ is given by

$$G_{N'}(\{\mathbf{r}_n\}) = k_B T \sum_n \frac{|\mathbf{r}_n - \mathbf{r}_{n-1}|^2}{2\ell_0{}^2} + \sum_{n < n'} U_{n,n'}.$$

We use this chain model only for sufficiently small stretching of the chain. The opposite case of chains so taut as to resist further stretching through molecular binding energy, rather than through entropic elasticity, requires more detailed, chain-specific, considerations.

We model the globular proteoglycans as particles at positions $\mathbf{X}_m$ (with $m = 1, \ldots, M$), interacting with one another *via* a two-body potential. We assume the effects of the inherent anisotropy of these molecules to be negligible, and thus model their interactions *via* a central-force potential $V$. Furthermore, we approximate $V$ as a short-range, hard-sphere potential of radius $a$, to account for the mutual impenetrability of the globules, which in essence holds at the relevant energy scales:

$$V_{m,m'} \equiv V_{\text{hs}}(|\mathbf{X}_m - \mathbf{X}_{m'}|) = \begin{cases} 0, & \text{for } |\mathbf{X}_m - \mathbf{X}_{m'}| > 2a; \\ \infty, & \text{otherwise.} \end{cases}$$

The partition function for one aggregate factors in all configurations of the chain segments $\{\mathbf{r}_n\}$ and of the hard spheres $\{\mathbf{X}_m\}$, as well as the locations $\{\sigma_m\}$ of the binding sites of the $m$th hard sphere along the chain backbone (which we may take to be ordered, *i.e.*, $\sigma_m < \sigma_{m+1}$ for all $m$). Each $\sigma_m$ may take on values in $\{1, \ldots, n\}$, and is associated with the binding constraint $\mathbf{X}_m = \mathbf{r}_{\sigma_m}$. Thus, in terms of the chemical potential $\mu$ of the suspended globules and the binding energy $E_b$, the Boltzmann factor for creating a globule bound to a monomer of the chain is $e^{\beta(\mu + E_b)}$.[24] For an aggregate formed in a sufficiently dilute suspension of globules, $\mu$ is related to the density of unbound globules $\rho_u$ *via* the expression $\mu = k_B T \ln(\rho_u a^3)$, obtained from the analogous expression for the chemical potential of an ideal gas. In this (grand canonical) ensemble (with respect to the globules), the partition function of an aggregate may be expressed as

$$Z = \sum_M e^{\beta(\mu + E_b)M} Z_M, \tag{3}$$

where $Z_M$ is the partition function for an aggregate with a fixed number $M$ of bound globules (*i.e.*, in the canonical ensemble) and is given by

$$Z_M = \sum_{\{\sigma_m\}}' \int \left[ \prod_{n=1}^{N'} d\mathbf{r}_n \right] \left[ \prod_{m=1}^{M} d\mathbf{X}_m \right] \left( \prod_{m < m'} e^{-\beta V_{m,m'}} \right) e^{-\beta G(\{\mathbf{r}_n\})}$$
$$\times \prod_{m=1}^{M} \delta^{(3)}(\mathbf{X}_m - \mathbf{r}_{\sigma_m}), \tag{4}$$

in which $\sum'$ is the sum over all ordered binding locations $\{\sigma_m\}$. Thus, this partition function considers all of the configurations of a single aggregate, and assigns each configuration the appropriate weight. We use the assumptions that $N' \gg 1$ and $M \gg 1$, so we need not consider polymer end effects. Thus, we replace $N'$ by $N = \sigma_M - \sigma_1 + 1$, and chose the index $n$ to run from 1 to $N$. As a consequence of the assumption that $M \gg 1$, the sum in eqn (3) can be approximated by an integral, with the result that

$$Z \approx \int dM e^{\beta(\mu + E_b)M} Z_M. \tag{5}$$

To achieve further progress, we make the following assumptions, which are mathematical consequences of the physical ranges of the parameters necessary for the aggregates to form:

(1) Only "nearest-neighbor" globule–globule interactions matter. Thus, the globule at $\mathbf{X}_m$ only interacts with those at $\mathbf{X}_{m\pm 1}$.

(2) $|\mathbf{r}_{\sigma_m} - \mathbf{r}_{\sigma_{m'}}| - 2a \ll 2a$ for all such pairs, indicating that the interacting hard-core globules almost touch each other.

(3) $|\sigma_n - \sigma_{n-1}| = N/M \gg 1$, indicating that the bound globules are evenly spaced along the contour of the polymer chain.

The limits of validity of these assumptions, and the corrections that would arise from relaxing them, are explored in the appendix. As a consequence of Assumption 1, the partition function depends on the coordinates of the globules only through the combinations $\mathbf{X}_m - \mathbf{X}_{m-1}$ for $m = 2 \ldots M$. Furthermore, using Assumption 2, the inter-globule interaction simplifies *via* $e^{-\beta V_{m,m-1}} \to \delta(|\mathbf{X}_m - \mathbf{X}_{m-1}| - 2a)$, which implements the notion that all pairs of adjacent, mutually impenetrable globules are strongly tethered to one another. Finally, using Assumption 3, the partition function decomposes into that of $M$ subchains, with each subchain being composed of $N/M$ segments and each being stretched to a distance $2a$ (and where $Z_1$ is the partition function for one subchain). Thus, we have:

$$Z_M \approx V(4\pi)^M (Z_1)^M = V(4\pi)^M \left( \int \left[ \prod_{n=2}^{N/M} d\mathbf{r}_n \right] \right.$$
$$\left. \times \delta(|\mathbf{r}_1 - \mathbf{r}_{N/M}| - 2a) e^{-\beta G_{N/M}(\{r_n\})} \right)^M, \tag{6}$$

where $V$ is the volume of the entire system and the factor $(4\pi)^M$ results from the orientational degrees of freedom of the subchains, *i.e.*, the freedom of the spheres to relocate around one another.

## A  A single ideal chain

For an ideal chain, for which $U = 0$, the free energy $F_{\text{id}} = -k_B T \ln Z_1$ of a single stretched chain has the form $F_{\text{id}} = \frac{3}{2} \frac{(2a)^2 k_B T}{(N/M)\ell_0{}^2}$,[25] which from eqn (6) results in the scaling,

$$Z_{M;\text{id}} \sim e^{-\frac{3}{2} \frac{(2a)^2 M^2}{N \ell_0{}^2}}. \tag{7}$$

To evaluate the mean, variance, and relative width of this (Gaussian) distribution, we substitute $Z_{M;\text{id}}$ into eqn (5) to

obtain the results

$$\langle M \rangle = \frac{1}{3} \beta (E_b + \mu) \frac{R_g^2}{(2a)^2}, \tag{8}$$

$$\langle (\delta M)^2 \rangle = \frac{1}{3} \frac{R_g^2}{(2a)^2}, \tag{9}$$

$$\frac{\langle (\delta M)^2 \rangle}{\langle M \rangle^2} = \frac{1}{3} [\beta (E_b + \mu)]^{-2} \frac{(2a)^2}{R_g^2} \sim \left( \frac{R_g}{a \langle M \rangle} \right)^2, \tag{10}$$

where for an ideal chain $R_g \equiv \sqrt{N} \ell_0$. Expression (10) indicates that the relative width of the distribution is small: $\frac{\langle (\delta M)^2 \rangle}{\langle M \rangle^2} \ll 1$, which follows because $R_g \ll \langle M \rangle a$, *i.e.*, because the bound globules swell the chain.

## B A single self-avoiding chain

The entropic elasticity of self-avoiding chains has two regimes: a harmonic regime, in which the chain is stretched much *less* than the radius of gyration; and an anharmonic regime, in which the chain is stretched much *more* than the radius of gyration. In the harmonic regime, $(N/M)^\nu \ell_0 \gg a$, *i.e.*, the radius of each "blob"[26] comprising the stretched chain is much greater than the size of the globular particle. In this regime, the globules all fit inside the blobs of the polymer chain and, therefore, the aggregate is only slightly swollen by the globules. By contrast, physical hyaluronan–proteoglycan aggregates are significantly larger than isolated chains, and therefore correspond to the anharmonic regime.

In this anharmonic regime, for which $(N/M)^\nu \ell_0 a$ and the chain is highly stretched,[25] we have from eqn (6)

$$Z_{M;sa} \sim e^{-C \left( \frac{2aM}{R_g} \right)^\delta}, \tag{11}$$

where the radius of gyration of the globule-free chain is $R_g \sim N^\nu \ell_0$ (for self-avoiding chains, $\nu = 3/5$ according to Flory theory), $\delta \equiv (1 - \nu)^{-1} = 5/2$,[25] and $C$ is a constant. Note that $Z_{M;sa}$ becomes $Z_{M;id}$ if the ideal chain values $\nu = \frac{1}{2}$ and $C = \frac{3}{2}$ are substituted into eqn (11). Next, the use of the method of steepest descent[27] gives the following expressions for the mean, variance, and relative width of the size distribution of aggregates for the case of self-avoiding polymer chains:

$$\langle M \rangle \approx \left( \frac{\beta(\mu + E_b)}{C\delta} \right)^{\frac{1}{\delta-1}} \left( \frac{R_g}{2a} \right)^{\frac{\delta}{\delta-1}}$$

$$= \left( \frac{\beta(\mu + E_b)}{C\delta} \right)^{\frac{1}{\nu}-1} \left( \frac{R_g}{2a} \right)^{\frac{1}{\nu}} \tag{12}$$

$$= \left( \frac{\beta(\mu + E_b)}{C\delta} \right)^{\frac{1}{\nu}-1} N \left( \frac{\ell_0}{2a} \right)^{\frac{1}{\nu}}$$
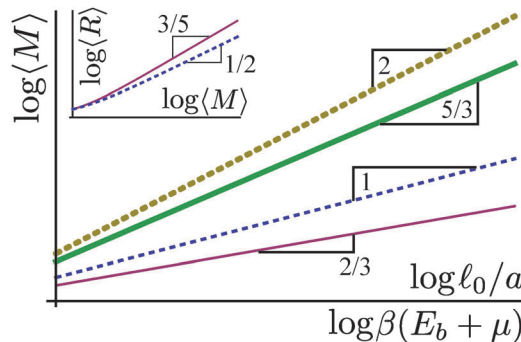


Fig. 2 Scaling laws [eqn (8) and (12)] for the number of globules $\langle M \rangle$ in an aggregate as a function of $\beta(\mu + E_b)$ (bottom, thin) and $\ell_0/a$ (top, thick), plotted on a log–log scale, with the numbers indicating the line slopes. The results from the ideal chain model are dashed, whereas those from the Flory theory of self-avoiding chains are solid. The inset shows the radius of gyration $\langle R \rangle$ of an aggregate as a function of $\langle M \rangle$: for sufficiently large aggregates, $\langle R \rangle$ scales as $\langle M \rangle^\nu$. The dashed line is the result for only nearest-neighbor globule–globule interactions, whereas the solid line is the result that includes interactions between all globule pairs.

$$\langle (\delta M)^2 \rangle \approx \frac{1}{C\delta(\delta - 1) (2a/R_g)^\delta \langle M \rangle^{\delta-2}}$$

$$= \frac{1}{C\delta(\delta - 1)} (\beta(\mu + E_b))^{\frac{1}{\nu}-2} \left( \frac{R_g}{2a} \right)^{\frac{1}{\nu}} \tag{13}$$

$$= \frac{1}{C\delta(\delta - 1)} (\beta(\mu + E_b))^{\frac{1}{\nu}-2} N \left( \frac{\ell_0}{2a} \right)^{\frac{1}{\nu}}$$

$$\frac{\langle (\delta M)^2 \rangle}{\langle M \rangle^2} \sim (\beta(\mu + E_b))^{-\frac{1}{\nu}} \left( \frac{R_g}{2a} \right)^{-\frac{1}{\nu}} \sim M^{-1} (\beta(\mu + E_b))^{-1} \ll 1 \tag{14}$$

In particular, if the Flory result $\nu = 3/5$ is inserted, these expressions simplify to eqn (2) (Fig. 2).

# III Many aggregates in suspension

To study the structure and thermodynamics of aggregates in suspension, we first consider the configurations of a single isolated aggregate, and then include the effects of inter-aggregate interaction. Thus, let us focus on an aggregate having $M$ bound globules, and consider the configurations of the aggregate that dominate $Z_M$. Under assumptions 1, 2, and 3, these configurations form a family characterized by the location of the center of mass of the aggregate, as well as the values for the $\sim 2M$ angular coordinates of $X_m - X_{m-1}$. Thus, the configuration space is identical to that of a freely-jointed polymer chain composed of $M$ segments of segment length $2a$. As a result, the characteristic linear size of a single aggregate is approximately $2a\sqrt{M}$.

However, Assumption 1 is too strong to describe a physical aggregate. Thus, we relax this assumption by allowing for interactions between arbitrary pairs of globules (which are labeled $m$ and $m'$). There are then two notable effects of these

additional interactions. First, for small $|m - m'|$, the interactions limit the ranges of the bond angles. For example, as a consequence of the fact that two next-nearest-neighbor spheres cannot overlap, the angle between globules $m$, $m + 1$, and $m + 2$ must lie within the range $\pi/3$ to $5\pi/3$. Such effects quickly diminish as $|m - m'|$ increases. Macroscopically, these interactions have the effect of stiffening the chain at fixed contour length. We take this effect into account by representing the aggregates as a freely-jointed chains having a segment length on the order of, but somewhat greater than, $2a$. Second, the avoidance between distant globules swells the aggregate, which, therefore, we represent as a self-avoiding chain. This effect is analogous to that of the swelling of a polymer chain in good solvent, with the consequent result for the aggregate size: $\langle R \rangle \sim 2aM^\nu$.[28,29] The size distribution of the aggregate, given by eqn (8) and (12)–(14), is not affected by these interactions. This is a consequence of the fact that the number $2M$ of orientational degrees of freedom of the effective aggregate chain is small compared to the total number $3N$ of chain degrees of freedom.

We now consider many aggregates in suspension, and explore their thermodynamics in three regimes—dilute, semidilute, and dense. As we shall see: (1) For sufficiently dilute suspensions, the properties of the individual aggregate should be largely independent of the aggregate concentration, with only small corrections due to aggregate–aggregate interactions. This results from a separation of energy scales, as the entropy of stretching is much larger than the aggregate-concentration-dependent free energy. On the other hand, (2) for sufficiently dense suspensions, the suspension behaves essentially as a dense suspension of globules, with only small corrections due to the presence of the polymer chains. This results from the fact that the concentration-dependent free energy is much larger than the entropy of chain stretching. (3) The cross-over between these two regimes occurs at densities sufficiently large that the suspension is described as dense rather than dilute or semidilute.

These results follow by considering a suspension of aggregates, and introducing the additional thermodynamic quantities $\rho$ (the aggregate number density), $V$ (the system volume), and $P$ (the number of aggregates). For convenience, we choose to work at fixed $\rho \equiv P/V$. We remark that the suspension is considered dilute if the characteristic size of the aggregate is much smaller than the inter-aggregate spacing, i.e., if $(M^\nu a) \ll (V/P)^{1/3} = \rho^{-1/3}$; see Fig. 3. For a dilute aggregate suspension, we may estimate the osmotic pressure $\Pi$ using the virial expansion as

$$\Pi = \rho T \left( 1 + \frac{B_2}{v_0} \rho v_0 + \cdots \right), \qquad (15)$$

where $v_0 \equiv \frac{4}{3}\pi (M^\nu a)^3$ is the volume of a single aggregate, $\rho v_0 \ll 1$ and $B_2/v_0$ is a numerical virial coefficient (which for a hard-sphere fluid has the value 4). Note that the first term on the right-hand side of eqn (15) is independent of aggregate size; a weak dependence on aggregate size results from the (small) second term. Thus, in the dilute regime, the total Helmholtz free-energy density of the suspension $f$ is given by
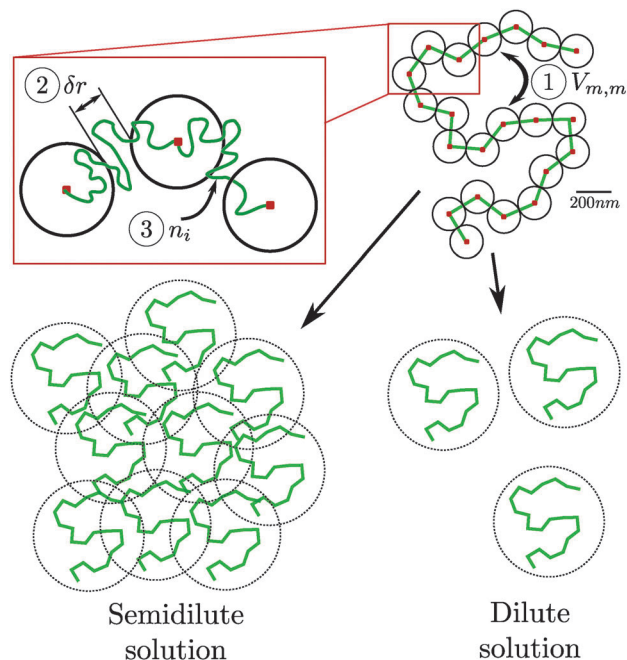
$$f = \Pi - \rho T \ln Z_M, \qquad (16)$$



Fig. 3 The structure of aggregates in suspension. The aggregates behave as self-avoiding polymer chains of segment length $2a$, composed of $M$ segments. In the dilute regime, the aggregates do not overlap, as shown on the bottom right. In the semi-dilute regime, as shown on the bottom left, the aggregates overlap at small globule packing fractions. In the dense regime (not shown), the globule packing fraction approaches 1. The assumptions are indicated using the circled numbers 1, 2, and 3.

which is dominated by the large chain-stretching term $-\ln Z_M \sim -(aM/R_g)^\delta \gg 1$. The second term in the virial expansion provides only a small correction to the result, shown in eqn (12), for $\langle M \rangle$, associated with deswelling of each aggregate.

We now turn to the semidilute regime, in which the suspended aggregates overlap, although the spacing between individual globules is large, relative to their size; see Fig. 3. This occurs for the density range $v_0^{-1} \ll \rho \ll a^{-3}$. By analogy with the ideal chains, a mean-field approximation for the properties of semi-dilute suspensions yields an expression for the free-energy density given by[30]

$$\beta f \sim \frac{1}{2}a^3 (M\rho)^2 + \rho C (2aM/R_g)^2. \qquad (17)$$

As $a^3\rho \ll 1$, the two terms are only comparable when $(2a/R_g)^2 \ll 1$, which may exist for some range of parameters. However, for the physical parameters of HA–PG aggregates, $a/R_g$ is of order one. This means that for semi-dilute suspensions of HA–PG aggregates, the concentration of aggregates will not affect aggregate size.

The scaling laws resulting from this mean-field theory can be corrected to account for aggregate-density fluctuations in suspensions of self-avoiding aggregates. To do so, we introduce the stretching exponent $\delta$ (=5/2); then the expression for the free-energy density becomes (see ref. 30)

$$\beta f \sim \frac{1}{2}a^{-3} (a^3 M\rho)^{9/4} + \rho C (2aM/R_g)^{5/2}. \qquad (18)$$

These fluctuation corrections suggest that only for parameter values such that $(a/R_g)^{5/2}M^{1/4} \gg (a^3\rho)^{5/4}$ would the aggregate concentration not affect aggregate size.

For solutions dense in globules, the osmotic pressure of the hard-sphere globules can dominate over the chain-stretching entropy. In that case, the chain degrees of freedom provide small corrections to a free energy that is essentially the same as the free energy of hard spheres of radius $a$. That is, sufficiently dense suspensions behave as dense suspensions of globules, with only small corrections arising from the chain binding.

## IV Discussion and conclusions

We conclude by first discussing the implications of the model developed in this work to hyaluronan–proteoglycan aggregates, and then discussing the implications of this model to other materials, including synthetic ones. For hyaluronan–proteoglycan aggregates, we have obtained the dependence of (i) the number of bound proteoglycans per aggregate and (ii) the size of the aggregate on model parameters such as the temperature, binding energy, size, and chemical potential of the globules, as well as the contour and persistence lengths of the hyaluronan chain. These predictions can, in principle, be experimentally tested in both biological and synthetic systems. To clarify these potential tests, we revisit an important assumption of the model, *viz.* the assumption that the polymer chain is flexible. As mentioned above, this assumption holds as long as the polymer chain segment between two neighboring bound globules is much longer than the polymer persistence length. This, therefore, sets limits on the number of bound globules in an aggregate for which the model remains applicable. We conjecture that for aggregates outside this limit, the binding of additional globules is suppressed due to the bending stiffness of the polymer chain. To test the scaling laws predicted by the model considered here within a single experiment, one may vary the chemical potential of the globules or, equivalently, the concentration of free globules in solution. Other parameters, such as the length-distribution of synthetic hyaluronan and the size of the aggrecan globules may also be varied from experiment to experiment. This model also leads us to obtain expressions for the free energy of a suspension of many aggregates for several regimes, such as dilute (with respect to the aggregate concentration) or semi-dilute (and still dilute with respect to the globules). We thus predict—in terms of the aggregate parameters—the osmotic pressure and compressibility of the suspension and, in the semi-dilute regime, the presence of a correlation length (observable *via* scattering), as well as the 9/4 scaling of the osmotic pressure with the aggregate density. These predictions, too, can be tested in experiments on either synthetic or biological hyaluronan–proteoglycan systems, and can be used to deduce materials parameters of these suspensions. Thus, the thermodynamic results obtained here may lead to advances in the understanding of neuroplasticity as well as the relationship between joint mechanics and materials properties of cartilage.

The beads-on-a-string model, which we have presented here, is also sufficiently general to describe materials beyond the example of hyaluronan–proteoglycan aggregates on which we have focused. This model combines two basic models of statistical mechanics: the hard-sphere model, which describes suspensions, and the random walk model, which describes polymer chains. It thus has the potential to describe any material composed of a suspension of large globular particles that have specific binding sites through which the globules bind to flexible polymer chains. A possible candidate for such a material, designed to realize the theory presented here, may be envisioned as a synthetic system composed of colloidal particles and polymer chains. For this system to aggregate according to the model we have considered, the colloidal particles must be functionalized with a single binding site that has a strong affinity to be bound to a polymer chain, which could be accomplished, *e.g.*, with complementary strands of DNA.[31] The chains must be sufficiently long to accommodate many bound colloids, and the binding energy must be sufficiently strong to promote aggregate formation, but not so strong as to lead to irreversible binding. Within this artificial system, it may be possible to control such parameters as the binding energy (*via* the DNA strand length and sequence), the temperature, and the colloid concentration and size – either during the preparation phase of the experiment or during the experiment itself. A suspension composed of these microscopic aggregates would be a biologically inspired fluid metamaterial, and have the remarkable property that the size of the aggregates is reversibly controllable. Therefore, the model presented in this work could guide the creation of suspensions that have finely tunable viscoelasticity as well as exhibit reversible sedimentation. By studying the properties of macroscopic suspensions of these colloid-polymer aggregates, one may gain further understanding of the materials properties that biology has realized as a result of natural selection, as well as use this understanding for materials engineering.

## Appendix A: revisiting initial assumptions

We revisit the derivation of eqn (8) and (12)–(14), justifying assumptions 2 and 3 within a mathematical framework and exploring the corrections to the aggregate partition function that result from relaxing these assumptions.

First, let us examine a generic pair-interaction potential between globules $V(r)$ and consider the distance between two "nearest-neighbor" globules $r \equiv |\mathbf{X}_m - \mathbf{X}_{m-1}|$ (see Fig. 3) and the statistical distribution of that distance, *i.e.*, $e^{-(\beta V(r)+F(r))M}$, where $F(r)$ is the free energy of chain stretching. This distribution is maximized near $V'(r) = -F'(r)$, and is sharply peaked for sufficiently large $M$. If we consider a sequence of potentials that approach $V_{hc}$, the solution to this equation must approach $r = 2a$, showing that indeed the distribution is dominated by small inter-globule separations.

For the hard-sphere potential, the validity of Assumption 2 may be tested directly by calculating $\langle r - 2a \rangle$ and higher

moments of the distributions $Z_{M;\mathrm{id}}$ or $Z_{M;\mathrm{sa}}$. By definition,

$$\langle r^A \rangle = K^{-1} \int_{2a}^{\infty} e^{-C\left(\frac{rM}{R_g}\right)^{\delta}} r^A \mathrm{d}r, \tag{A1}$$

where $K \equiv \int_{2a}^{\infty} e^{-C\left(\frac{rM}{R_g}\right)^{\delta}} \mathrm{d}r$ is the normalization. The ideal-chain case is obtained by taking $C = 3/2$, $\delta = 2$, in which case the moments of $r$ may be evaluated exactly, in terms of the Error function. For arbitrary values of $\delta$ in the range $\delta > 1$, the above expectation value can be evaluated using the relation

$$\int_{\frac{2aM}{R_g}}^{\infty} e^{-Cx^{\delta}} x^A \mathrm{d}x = \frac{1}{\delta C} \left(\frac{2aM}{R_g}\right)^{A+1-\delta} e^{-C\left(\frac{2aM}{R_g}\right)^{\delta}}$$
$$+ \frac{A+1-\delta}{\delta C} \int_{\frac{2aM}{R_g}}^{\infty} e^{-Cx^{\delta}} x^{A-\delta} \mathrm{d}x, \tag{A2}$$

which is obtained using integration by parts. On the right-hand side of this expression, the first term is much larger than the second term, and the expression can therefore be used to obtain an asymptotic expansion for the integral on the left-hand side. Iterating this relation to next order, we obtain (using $A = 0$ and 1),

$$\frac{\langle r - 2a \rangle}{2a} \approx \frac{1}{\delta C} \left(\frac{R_g}{2aM}\right)^{\delta} \ll 1. \tag{A3}$$

Using an additional iteration of eqn (A2) ($A = 0$, 1, and 2), we obtain

$$\frac{\langle (r - \langle r \rangle)^2 \rangle}{\langle r \rangle^2} \approx \frac{1}{\delta^2 C^2} \left(\frac{R_g}{2aM}\right)^{2\delta} \approx \left(\frac{\langle r - 2a \rangle}{2a}\right)^2 \ll 1.$$

Thus, assumption 2 is justified on the basis that the deviations of $r$ from the value $2a$ are small, both in terms of the expectation value of $r$, and fluctuations around this expectation value.

Finally, let us now relax Assumption 3 by considering the distribution of $n_i \equiv \sigma_n - \sigma_{n-1}$ (see Fig. 3). If each $n_i$ is considered as an identically distributed independent random variable, it follows from the central limit theorem that if the mean of the sum of $n_i$ is N, the mean of each $n_i$ is indeed $N/M$. To calculate the width of the distribution, we consider the ensemble of fixed $M$ and varying chain contour length $N$, determined by a conjugate variable $p_N$. Thus, the corresponding partition function is given by the sum over all possible arrangements of globules along the chain:

$$Z_{MN;\mathrm{sa}} \approx \sum_{\{n_i\}_1^M n_i=1}^{N} \delta_{N,\sum n_i} e^{-\sum_{n_i} C\left(\frac{2a}{\ell_0^2 n_i^{\nu}}\right)^{\delta}}. \tag{A4}$$

In the $p_N$ ensemble, after summing over the chain contour lengths $N$, this partition function has the form

$$Z_{Mp_N;\mathrm{sa}} \approx \sum_{\{n_i\} n_i=1}^{N} e^{-\sum_{n_i} C\left(\frac{2a}{\ell_0^2 n_i^{\nu}}\right)^{\delta}} e^{-\beta p_N \sum_{n_i} n_i}. \tag{A5}$$

For large $n_i$, each sum may be approximated by an integral, giving

$$Z_{Mp_N;\mathrm{sa}} \approx \left( \int_{-\infty}^{\infty} \mathrm{d}n_i e^{-C\left(\frac{2a}{\ell_0^2 n_i^{\nu}}\right)^{\delta} - \beta p_N n_i} \right)^M. \tag{A6}$$

Substituting $\langle n_i \rangle \approx N/M$ for all $i$, we find the "equation of state" to be

$$\beta p_N \sim \left(\frac{2aM}{\ell_0 N}\right)^{\delta}. \tag{A7}$$

By using this approach, we may also calculate the relative width of the distribution of $n_i$, obtaining,

$$\frac{\langle (n_i - N/M)^2 \rangle}{(N/M)^2} \sim \left(\frac{\ell_0 N^{\nu}}{aM^{\nu}}\right)^{\delta} \ll 1, \tag{A8}$$

which reveals that, indeed, $n_i$ is narrowly distributed around the mean $N/M$. This narrowness holds for self-avoiding chains ($\nu = 3/5$) as well as for ideal chains ($\nu = 1/2$), in both cases under the condition $\frac{\ell_0 N^{\nu}}{aM^{\nu}} \ll 1$, i.e., in the physically relevant regime for the aggregates.

## References

1  P. Jenkins and M. Snowden, *Adv. Colloid Interface Sci.*, 1996, **68**, 57.
2  R. Tuinier, J. Rieger and C. G. de Kruif, *Adv. Colloid Interface Sci.*, 2003, **103**, 1.
3  E. Dickinson and L. Eriksson, *Adv. Colloid Interface Sci.*, 1991, **34**, 1.
4  B. Cabane, *J. Phys. Chem.*, 1977, **81**, 1639.
5  E. D. Goddard, *Colloids Surf.*, 1986, **19**, 255–329.
6  P. Hansson and B. Lindman, *Curr. Opin. Colloid Interface Sci.*, 1996, **1**, 604.
7  S. Alexander, *J. Phys.*, 1977, **38**, 977.
8  P. Pincus, C. J. Sandroff and T. A. Witten, *J. Phys.*, 1984, **45**, 725.
9  M. K. Cowman and S. Matsuoka, *Carbohydr. Res.*, 2005, **340**, 791.
10  T. Wight, B. Toole and V. Hascall, *Hyaluronan and the Aggregating Proteoglycans*, Springer, Berlin Heidelberg, 2011, ch. 5, pp. 147–195.
11  B. P. Toole, *Nat. Rev. Cancer*, 2004, **4**, 528.
12  Y. Yamaguchi, *Cell. Mol. Life Sci.*, 2000, **57**, 276.
13  H. Boehm, *et al.*, *Soft Matter*, 2009, **5**, 4331.
14  L. McLane, *et al.*, *Biophys. J.*, 2013, **104**, 986.
15  T. Hardingham, *Osteoarthr. Cartil.*, 1998, **6**(Suppl. A), 3.
16  C. Ricciardelli, *et al.*, *J. Biol. Chem.*, 2007, **282**, 10814.

17 S. P. Evanko, J. C. Angello and T. N. Wight, *Arterioscler., Thromb., Vasc. Biol.*, 1999, **19**, 1004.

18 D. Wang and J. Fawcett, *Cell Tissue Res.*, 2012, **349**, 147.

19 L. Ng, *et al.*, *J. Struct. Biol.*, 2003, **143**, 242.

20 L. Ng, *et al.*, *Macromol. Symp.*, 2004, **214**, 1.

21 R. J. Nap and I. Szleifer, *Biophys. J.*, 2008, **95**, 4570.

22 T. E. Hardingham and H. Muir, *Biochem. J.*, 1974, **139**, 565.

23 M. Mörgelin, *et al.*, *Biochem. J.*, 1988, **253**, 175.

24 In addition to the energetic gain for binding, there will, in general, also be an entropic cost due to the loss of polymer configurations when the chain is in proximity to an impenetrable globule. This effect would result in a contribution to the binding free energy that is linear in the temperature $T$. We gratefully acknowledge an anonymous referee for this remark.

25 See, *e.g.*, ref. 32, Section I.4.

26 P. Pincus, *Macromolecules*, 1976, **9**, 386.

27 The steepest descent calculation relies on the approximation $\int \mathrm{d}x f(x) \mathrm{e}^{g(x)} \approx \sqrt{\dfrac{2\pi}{-g''(x_0)}} f(x_0)$, where $g'(x_0) = 0$ is a unique local maximum at the value $x_0$. This may be extended to calculate the second moment using $\left(\int \mathrm{d}x \mathrm{e}^{g(x)}\right)^{-1} \int \mathrm{d}x (x - x_0)^2 \mathrm{e}^{g(x)} \approx \dfrac{1}{-g''(x_0)}$.

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

29 See, *e.g.*, ref. 32, Section I.3.

30 See, *e.g.*, ref. 32, Sections III.1 and III.2.

31 C. A. Mirkin, *et al.*, *Nature*, 1996, **382**, 607.

32 P. G. de Gennes, *Scaling Concepts in Polymer Physics*, Cornell University Press, 1979.