Depletion-induced instability in protein-DNA mixtures: Influence of protein charge and size

Renko de Vries[a)]
Laboratory of Physical Chemistry and Colloid Science, Wageningen University, P.O. Box 8038, 6700 EK Wageningen, The Netherlands

(Received 22 March 2006; accepted 10 May 2006; published online 6 July 2006)

While there is abundant experimental and theoretical work on polymer-induced DNA condensation, it is still unclear whether globular proteins can condense linear DNA or not. We develop a simple analytical approximation for the depletion attraction between rodlike segments of semiflexible polyelectrolytes such as DNA, induced by nonbinding globular proteins. Monte Carlo simulations show that the approximation works well up to protein volume fractions of at least 20%. From the expression for the depletion attraction we estimate instability thresholds by computing the effective virial coefficient of DNA segments in protein solutions. While globular proteins are found to be much poorer depletion agents than flexible polymers, it should be possible to condense linear DNA with small highly charged globular proteins, at relatively low ionic strengths. © *2006 American Institute of Physics*. [DOI: 10.1063/1.2209683]

INTRODUCTION

Phase separation induced by depletion interactions is a ubiquitous phenomenon in mixtures of macromolecules, colloids, surfactants, etc. Presumably because the main contribution to the excluded volume is due to purely steric interactions, most theories and simulations focus on model systems interacting through steric repulsion only, such as mixtures of hard spheres and hard rods[1–3] or mixtures of athermal flexible polymer chains with small hard spheres.[4] However, interaction effects are far from negligible in most systems of practical interest such as, for example, biopolymer mixtures.

A case in point is the classic ψ condensation of DNA (polymer and salt induced) which is driven by depletion interactions but, as the name suggest, strongly depends on ionic strength. Since the initial work of Lerman,[5] a substantial amount of work has been done on DNA ψ condensation, both experimentally[6–12] and theoretically.[10,13–18] For all of these studies, an important motivation is the possibility that globular proteins could act in a similar way as the polymers, to condense DNA in living cells.

Suprisingly, there is very little experimental work on DNA condensation by non-DNA-binding globular proteins and on its dependence on protein size and charge, ionic strength, etc. In fact, we are not aware of any reports of DNA condensation by non-DNA-binding globular proteins alone. Experiments of Murphy and Zimmerman[19] show that non-DNA-binding fractions of *E. coli* cell extracts by themselves are not sufficient to induce condensation of restriction fragments of λ DNA. Similarly, dense solutions of bovine serum albumin did not condense λ DNA restriction fragments.

On theoretical grounds, Castelnovo and Gelbart[18] also conclude that globular proteins are much less efficient at condensing DNA than inert flexible polymers such as polyethylene oxide (PEO). Assuming that the DNA and proteins behave as, respectively, hard cylinders and hard spheres (as far as their interactions are concerned), they predict that only "proteins" with unrealistically small radii (less than about 1.5 nm) could possibly condense DNA.

On the other hand, taking into account details of the electrostatic protein-DNA interactions, Odijk[20] argues theoretically that for supercoiled DNA, such as the genomic DNA of bacteria, depletion interactions with globular proteins are strong enough to cause phase separation into a DNA rich "nucleoid" and a protein rich "cytosolic" phase.

Extending the approach of Odijk, we here study the depletion of globular proteins around semiflexible polyelectrolytes such as DNA, anionic polysaccharides, and protein filaments, taking into account details of the protein-protein and protein-polyelectrolyte interactions. We derive simple analytical approximations based on a virial expansion of the polyelectrolyte chemical potential. Using Monte Carlo simulations we show that the approach is accurate up to protein volume fractions of at least 20%.

In particular, we derive a simple analytical expression for the globular protein-induced depletion attraction between parallel rodlike segments of semiflexible polyelectrolytes. By calculating the effective virial coefficient of segments of the semiflexible polyelectrolytes in protein solutions, we derive estimates for when mixtures with nonbinding globular proteins become unstable. Like Castelnovo and Gelbart, we find that nonbinding globular proteins are much less efficient depletion agents than inert flexible polymers. Nevertheless, in experiments with linear DNA, instability could be observed for small proteins with a high charge, at relatively low ionic strengths.

ANALYTICAL APPROXIMATION

The polyelectrolyte is assumed to be rodlike at the scale of a single globular protein. Then, the thermodynamics of the

[a)]Electronic mail: renko.devries@wur.nl

protein-polyelectrolyte interactions can be characterized in terms of the Helmholtz energy $f_{\rm ins}$ (energy per unit length) of inserting rodlike polyelectrolyte into a suspension of globular proteins. The virial expansion of $f_{\rm ins}$ is essentially that of the chemical potential in a multicomponent system. For a single, fixed, rodlike polyelectrolyte molecule, up to second order in the protein number density ρ_p,

$$f_{\rm ins}^{(1)} = B_{pd}\rho_p + \tfrac{1}{2}C_{ppd}\rho_p^2 + \cdots . \tag{1}$$

Note that we use subscripts p and d, respectively, for the proteins and the semiflexible polyelectrolytes (d for DNA, the case that we mostly consider here). Virial coefficients are given in terms of the interaction potentials $u_{pd}(r)$ and $u_{pp}(r)$ for protein-polyelectrolyte and protein-protein interactions,

$$B_{pd} = \int_0^\infty 2\pi r dr[1 - \exp(-u_{pd}(r)/kT)], \tag{2}$$

$$C_{ppd} = \frac{1}{L}\int d\mathbf{r}_1 \int d\mathbf{r}_2[1 - \exp(-u_{pp}(|\mathbf{r}_1 - \mathbf{r}_2|)\ ] \times[1 - \exp(-u_{pd}(r_1)/kT)][1 - \exp(-u_{pd}(r_2)/kT)]. \tag{3}$$

Here kT is the thermal energy, $\mathbf{r}_1$ and $\mathbf{r}_2$ are the positions, respectively, of protein 1 and 2, and r_1 and r_2 are distances to the fixed rodlike polyelectrolyte segment of length L.

Numerical evaluation of C_{ppd} is not too difficult, but when desired one can also use a simple approximation. Assuming that the cross-virial coefficient C_{ppd} is dominated by configurations in which only one of the proteins overlaps with the rod, we find that the integral over the positions of this particle contributes a factor of order LB_{pd}. The integral over the position of the other particle then contributes a factor of order B_{pp}, where B_{pp} is the protein-protein second virial coefficient,

$$B_{pp} = \frac{1}{2}\int_0^\infty 4\pi r^2 dr[1 - \exp(-u_{pp}(r)/kT)]. \tag{4}$$

Hence, a rough estimate of the cross-virial coefficient C_{ppd} is

$$C_{ppd} \approx B_{pp}B_{pd}. \tag{5}$$

This approximation was tested for the interaction potentials and parameter values used in the computer simulations (see Table II, discussed in the next section). It is found that for protein second virial coefficients B_{pp} that are sufficiently large, the approximation works surprisingly well. On the other hand, for proteins close to their theta point, Eq. (5) is a rather poor approximation. But in this case the cross-virial coefficient C_{ppd} is small anyway, and the major contribution to the insertion free energy in Eq. (1) is due to the lowest order cross-virial B_{pd}. In short, Eq. (5) can be used as a first order approximation both for proteins far away from the theta point and for proteins close to the theta point.

Next consider the effective potential of interaction (or potential of mean force) $V(r)$ of rodlike parallel segments of semiflexible polyelectrolytes at a distance r (per unit length of polyelectrolyte). It has two contributions: the direct polyelectrolyte-polyelectrolyte potential of interaction $u_{dd}(r)$ and a protein-induced contribution $\Delta V(r)$,

$$V(r) = u_{dd}(r) + \Delta V(r). \tag{6}$$

The range $D_{\rm attr}$ of the latter approximately equals the distance between the polyelectrolyte strands beyond which proteins start entering the space between them. A first order estimate is simply

$$D_{\rm attr} \approx 2(a_p + a_d + \kappa^{-1}), \tag{7}$$

where κ^{-1} is the range of the electrostatic repulsion, the Debye length, which depends on the monovalent electrolyte concentration n_s (number density) according to $\kappa^2 = 8\pi l_B n_s$. The Bjerrum length $l_B = e^2/\varepsilon kT$, where e is the elementary charge and ε is the solvent permittivity.

The protein-induced interaction $\Delta V(r)$ can be computed from Helmholtz energies of inserting semiflexible polyelectrolyte in protein suspensions,

$$\Delta V(r) = f_{\rm ins}^{(2)}(r) - f_{\rm ins}^{(1)}, \tag{8}$$

where $f_{\rm ins}^{(2)}(r)$ is the Helmholtz energy (per unit length of polyelectrolyte) of inserting two parallel rodlike polyelectrolytes at a distance r. Inserting two completely overlapping rods, at $r=0$, costs approximately the same as inserting a single isolated rod,

$$f_{\rm ins}^{(2)}(0) \approx \tfrac{1}{2}f_{\rm ins}^{(1)}. \tag{9}$$

The functional form of the distance dependence can be approximated by computing the overlap of the volumes excluded to the proteins by the polyelectrolyte molecules. This gives involved expressions, but the overall dependence on distance is nearly linear. Beyond a distance $D_{\rm attr}$, given by Eq. (7), the protein-induced interaction vanishes,

$$\Delta V(r) \approx \begin{cases} -\tfrac{1}{2}f_{\rm ins}^{(1)}(1 - r/D_{\rm attr}), & r \leqslant D_{\rm attr} \\ 0, & r > D_{\rm attr}. \end{cases} \tag{10}$$

MONTE CARLO SIMULATIONS

We have tested some aspects of the above approximations using Monte Carlo simulations. Following Prinsen and Odijk,[21] we assume that protein-protein interactions are isotropic and consist of steric repulsion, a short-ranged attraction and longer-ranged electrostatic repulsion,

$$u_{pp}(r) = u_{pp}^{\rm st}(r) + u_{pp}^{\rm attr}(r) + u_{pp}^{\rm el}(r). \tag{11}$$

For the simulations we prefer to use continuous rather than the discontinuous potentials of interactions often used in analytical work. Electrostatic interactions are calculated using the Poisson-Boltzmann equation, in the weak overlap approximation,

$$\frac{u_{pp}^{\rm st}(r)}{kT} = \left(\frac{2a_p}{r}\right)^{12}, \tag{12}$$

$$u_{pp}^{(\rm attr)}(r) = -U_A \exp\left(-\tfrac{1}{2}(r - 2a_p)^2/(2a_p\delta_A)^2\right), \tag{13}$$

$$\frac{u_{pp}^{\mathrm{el}}(r)}{kT} = \alpha_p^2 l_B \frac{\exp(-\kappa r)}{r}. \tag{14}$$

The parameter δ_A determines the width of the attractive well; the charge parameter α_p is related to the protein charge and size and to the salt concentration as explained below.

Rodlike polyelectrolytes are modeled as a linear array of overlapping charged spheres with a radius a_d equal to the polyelectrolyte radius. The potential of interaction of a protein sphere with a polyelectrolyte sphere is

$$\frac{u_{pd}^{(s)}(r)}{kT} = \left(\frac{a_p + a_d}{r}\right)^{12} + l_B \alpha_p \alpha_d \frac{\exp(-\kappa r)}{r}. \tag{15}$$

The distance between the spheres is taken to be equal to the Bjerrum length l_B, the superscript (s) indicates that this is for a single polyelectrolyte sphere. Summing over the contributions of all polyelectrolyte spheres we get the potential of interaction of the entire rodlike polyelectrolyte with a protein sphere,

$$\frac{u_{pd}(r)}{kT} = \sum_{n=-\infty}^{\infty} u_{pd}^{(s)}(\sqrt{r^2 + n^2 l_B^2})/kT \approx \left(\frac{a_p + a_d}{r}\right)^{12} + \sqrt{2\pi}\alpha_p \alpha_d \frac{\exp(-\kappa r)}{(\kappa r)^{1/2}}. \tag{16}$$

The physical interpretation of the charge parameters is that of an equivalent point charge (α_p, in units of elementary charges e, for proteins) or equivalent dimensionless line charge (α_d, for polyelectrolytes) that in the Debye-Hückel approximation generates the asymptotic field predicted by the full Poisson-Boltzmann equation. They are given by

$$\alpha_p = \frac{Z_{\mathrm{eff}} \exp(\kappa a_p)}{(1 + \kappa a_p)}, \tag{17}$$

$$\alpha_d = \frac{X_{\mathrm{eff}}}{\kappa a_d K_1(\kappa a_d)}, \tag{18}$$

where Z_{eff} is an effective protein charge related to the true protein charge Z, and X_{eff} is an effective dimensionless linear charge density, related to the true dimensionless linear charge density of DNA, $X = l_B \nu$, where ν is the DNA linear charge density (number of elementary charges per unit length). In the Debye-Hückel approximation, valid for weakly charged spheres and cylinders, $Z_{\mathrm{eff}} \approx Z$ and $X_{\mathrm{eff}} \approx X$. Convient closed formulas for $\kappa a > 1$ in the nonlinear Poisson-Boltzmann approximation have recently been derived by Aubouy *et al.*[22] and are used here. For ionic strengths $n_s > 0.05M$, as considered here, to a fair approximation, we may set $Z_{\mathrm{eff}} \approx Z$ for the proteins.

Simulation method

Metropolis Monte Carlo simulations are performed in the canonical ensemble. Proteins and rodlike DNA are enclosed in a square box with sides l_{box}. Periodic boundary conditions are employed, interactions are truncated using the minimum image convention. Positions of the polyelectrolyte spheres that make up the rodlike polyelectrolyte are fixed during the simulations. Trial moves for the protein positions consist of simple translations. New protein positions are uniformly sampled in a cubic box with sides Δl, centered on the old protein position. The value of Δl is chosen such that the average acceptance of trial moves is close to 0.5.

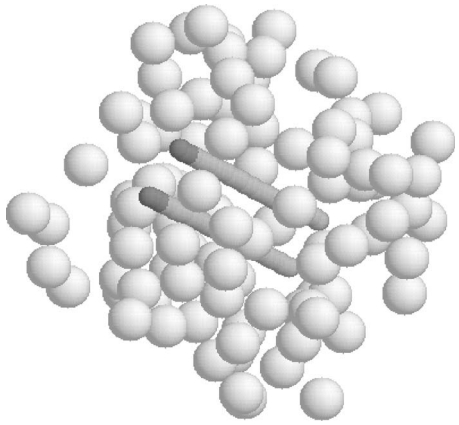

FIG. 1. Snapshot of a configuration from Monte Carlo simulations. This particular system has two finite sized rods (each consisting of 25 partially overlapping spheres) and 50 spherical particles. The dark parts of the rods are inserted in the Widom insertion moves for calculating the chemical potential.

Helmholtz energies of inserting polyelectrolytes into protein solutions are measured using the Widom insertion technique for measuring chemical potentials.[23] Widom insertions are attempted after every Monte Carlo (MC) cycle (trial moves per particle). In an insertion move, fixed rodlike polyelectrolyte is elongated by inserting N_{insert} spheres at one of the ends of the molecule. For the simulations with two parallel polyelectrolytes, N_{insert} is taken to be even, and both molecules are elongated by $N_{\mathrm{insert}}/2$ spheres, at the same end, as illustrated in Fig. 1. The Helmholtz energy of insertion is calculated in the usual way from the average value of the insertion probability.

Due to the presence of the polyelectrolytes, the average protein concentration in the simulation box is not the same as the bulk protein concentration, far away from the polyelectrolye molecules. Therefore, for each simulation we also measure the average (bulk) protein concentration far away from the polyelectrolytes.

Parameter values

Parameter values are listed in Table I and pertain to the case of DNA. The charge parameter for DNA is calculated assuming a bare linear charge density of $X = 4.2$ elementary

TABLE I. Parameter values used in the Monte Carlo simulations.

n_s (M)	0.15
κ^{-1} (nm)	0.77
l_B (nm)	0.71
a_d (nm)	1.2
a_p (nm)	2.5
α_d	7.0
α_p ($Z_{\mathrm{eff}}=5$)	30.3
α_p ($Z_{\mathrm{eff}}=15$)	90.8
U_A (kT)	2.68
δ_A	0.06

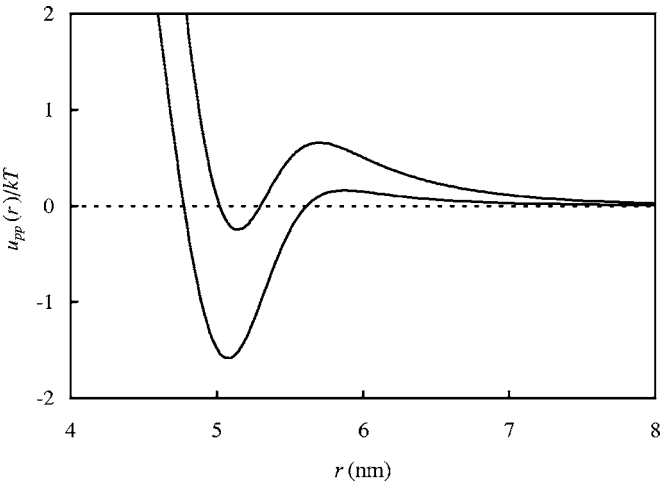

FIG. 2. Protein-protein interaction potentials $u_{pp}(r)$ used in the simulation. Bottom curve, $Z_{\text{eff}}=5$ ($B_{pp}\approx 0$), top curve, $Z_{\text{eff}}=15$.

charges per Bjerrum length (based on the density of phosphate groups along the DNA backbone, which gives a charge spacing along the axis of $0.17e/\text{nm}$). We consider a "typical" globular protein of size $a_p=2.5$ nm, for two values of the charge (or pH), at a physiological ionic strength of $n_s=0.15M$. Concerning the short-range attraction, Prinsen and Odijk[21] argue that it is the product $\delta_A \exp(U_A/kT)$ that is most strongly constrained by experimental values of protein-protein virial coefficients, rather than the interaction range itself, for which values anywhere in between 1.5 and 8 Å are consistent with experimental protein-protein virial coefficients for lysozyme. Interactions that are thought to contribute to the short range attraction are especially hydrogen bonds and hydrophobic interactions, both of which are rather short ranged indeed. Hence we set the range of the short-range attraction to a typical value of 3 Å, which gives $\delta_A=0.06$. We want to explore the whole range of protein solubilities from poorly soluble proteins (close to their theta point, $B_{pp}\approx 0$) to highly soluble (far away from their theta point). We choose two values for the protein charge: $Z\approx Z_{\text{eff}}=5$ and 15, representing the two extremes of a poorly soluble protein and a highly soluble protein. For the lowest value $Z\approx Z_{\text{eff}}=5$ to represent the poorly soluble protein, we have to choose the strength of the attraction such that the protein-protein second virial coefficient B_{pp} vanishes. This gives $U_A=2.68$. At the higher value of $Z\approx Z_{\text{eff}}=15$ the attraction makes only a small contribution to the virial coefficient and is instead dominated by steric and electrostatic repulsions, more characteristic of a highly soluble protein. A plot of the protein-protein potential of interaction for the two cases is shown in Fig. 2.

The size of the simulation box is $l_{\text{box}}=50$ nm. The number of proteins is varied between 50 and 350. Rodlike DNA molecules consist of 25 overlapping spheres, which gives a rod length of $25l_B\approx 17.8$ nm. We have performed simulations both for single rods and for pairs of parallel rods on a distance r. For determining Helmholtz insertion energies, rods are elongated by $N_{\text{insert}}=4$ spheres, equivalent to a length of 2.81 nm. Each simulation run consists of 5×10^5 MC cycles on equilibrated initial configurations.

TABLE II. Virial coefficients for the parameters values used in the Monte Carlo simulations. $C_{ppd}^{(a)}$ is the cross-virial coefficient according to the approximation of Eq. (5).

	$Z_{\text{eff}}=5$	$Z_{\text{eff}}=15$
B_{pp} (nm^3)	0.95	387.7
B_{pd} (nm^2)	72.0	94.2
C_{ppd} (nm^5)	3703±29	39 433±11
$C_{ppd}^{(a)}$ (nm^5)	68.4	35 673.54

For comparison with the analytical approximations, we have numerically evaluated the virial coefficients for the potentials of interaction used in the simulations. These are listed in Table II. Third order virial coefficients C_{ppd} were evaluated using simple sampling Monte Carlo integration using Sobol quasirandom number sequences.[24] For comparison, Table II also gives approximate values for C_{ppd} computed with Eq. (5).

Comparison of analytical approximation and simulation results

Results for the Helmholtz energy of inserting a single rodlike DNA molecule are shown in Fig. 3 For the protein close to its theta point ($B_{pp}\approx 0$), the insertion energy is nearly linear in the protein concentration over the entire range that we have probed. The virial expansion, Eq. (1), fits the data reasonably well up to the highest protein concentration, which corresponds to a protein volume fraction of around 20%.

Next consider the results for the Helmholtz energy of inserting parallel rodlike DNA molecules at a distance r. By subtracting the Helmholtz energy of inserting isolated rods, we obtain the protein-induced potential of interaction $\Delta V(r)$. Figure 4 shows some of the data for the protein close to its theta point ($Z_{\text{eff}}=5$). Since we are subtracting two chemical potentials to obtain a small difference, the errors are relatively large. For this case (with $B_{pp}=0$ and $C_{ppd}\approx 0$), within

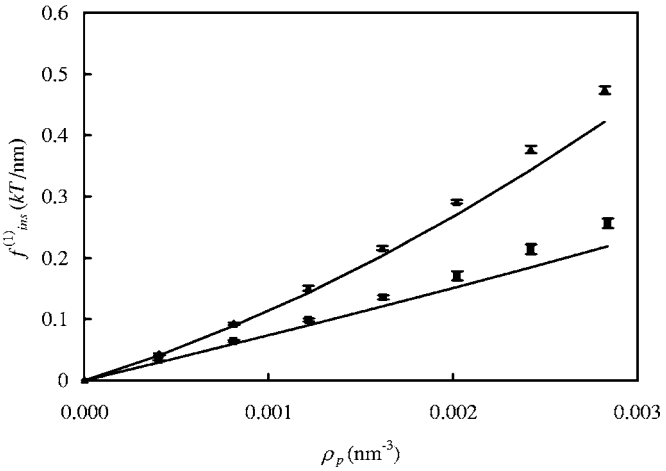

FIG. 3. Insertion free energy $f_{\text{ins}}^{(1)}$ vs protein density ρ_p. Bottom curve, protein charge, $Z_{\text{eff}}=5$ ($B_{pp}\approx 0$), top curve, $Z_{\text{eff}}=15$. Lines are the virial expansion Eq. (1), as discussed in the text.

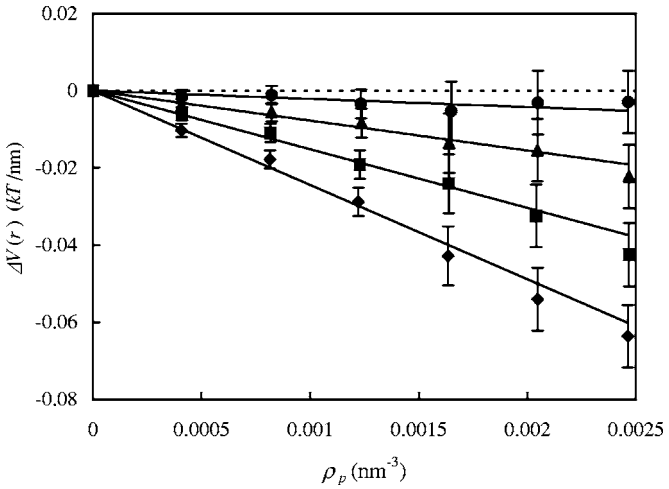

FIG. 4. Depletion attraction $\Delta V(r)$ (per unit length) as a function of protein density ρ_p for $Z_{\text{eff}}=5$ and various values of the spacing r of the parallel rodlike polyelectrolytes. From top to bottom: $r=8.5$, 7, 5, and 3 nm.

the error, the insertion energies are linear in the concentration, with a strong dependence on the distance r between the parallel rods.

Figure 5 shows the depletion attraction $\Delta V(r)$ scaled by the protein concentration ρ_p for the case $Z_{\text{eff}}=5$. For distances beyond approximately 9 nm, the protein-induced interaction vanishes. This is close to the analytical estimate Eq. (7), which gives $D_{\text{attr}}=8.9$ nm.

The simple approximation for $\Delta V(r)$, Eq. (10), fits the simulation data within the error. Similar agreement between theory and simulations is obtained for the protein with $Z_{\text{eff}}=15$. Clearly, the simple approximation Eq. (10) provides a reasonable first order approximation for the protein-induced potential of interaction, for protein volumes at least up to about 20%.

INSTABILITY OF LINEAR DNA

For the case of DNA ψ condensation it is well known that there is essentially no flexible polymer inside the condensates, except if the polymers are of very low molecular weight.[12] This allows for convenient analytical approximations in calculating phase boundaries.[17] However, for globular proteins, especially small ones, it is uncertain whether all proteins will be excluded, when condensates form. A full theory for the phase behavior of mixtures of globular proteins and semiflexible polyelectrolytes would therefore require that we deal with the many-body interactions in dense, possibly liquid-crystalline protein-semiflexible polyelectrolye mixtures.

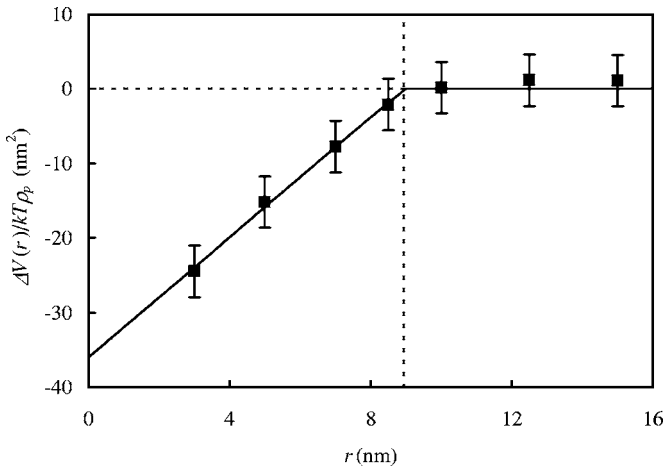

FIG. 5. Depletion attraction (scaled with protein density) as a function of the spacing r of the parallel rodlike polyelectrolytes, for $Z_{\text{eff}}=5$. The solid line is the simple approximation, Eq. (10).

This is beyond the scope of the present paper. Instead, we give simple estimates for when the mixtures become unstable by calculating the second virial coefficient B_{eff} of Kuhn-length segments of semiflexible polyelectrolytes immersed in protein solutions. When this turns negative, mixtures may be expected to be unstable against phase separation. Higher order virial coefficients may turn negative even before this threshold,[25,26] such that our estimate is a conservative one.

Note that there are certain limitations to this approach: we can only demonstrate the existence of an instability, but cannot tell anything about the nature of the condensed phase. In particular, we cannot tell whether the condensed phase will be isotropic, nematic, or hexagonal, or whether it will still contain proteins or not. We assume that the DNA concentration is low enough, such that there is only a weak concentration dependence of the condensation threshold, through the translational entropy of DNA molecules (especially important for short DNA fragments). This weak concentration dependence is neglected here, by assuming a threshold $B_{\text{eff}}=0$ irrespective of the number density of DNA molecules.

We consider two rodlike segments of the semiflexible polyelectrolytes, each with length $l_K=2P$, where P is the persistence length of the semiflexible polyelectrolytes. For DNA, we use $l_K=100$ nm. According to Eq. (6), the total interaction of parallel rodlike polyelectrolytes has contributions both due to the direct interactions and due to the depletion of globular proteins: $V(r)=u_{dd}(r)+\Delta V(r)$. For the direct electrostatic polyelectrolyte-polyelectrolyte repulsion we use the weak overlap approximation that we have also used for the electrostatic protein-protein and protein-polyelectrolyte interactions. Per unit length of parallel polyelectrolyte,

$$u_{dd}(r) \approx \sqrt{2\pi}\alpha_d^2 l_B^{-1}\frac{\exp(-\kappa r)}{(\kappa r)^{1/2}}, \quad r > 2a_d. \tag{19}$$

The protein-induced contribution $\Delta V(r)$ is approximated according to Eqs. (1), (5), and (10). Figure 6 shows the total potential of interaction for two parallel DNA segments of length l_K as a function of their separation r in the presence of increasing volume fractions of globular proteins of radius $a_p=2.5$ nm and charge $Z=15$, at an ionic strength of $n_s=0.15M$.

The effective virial coefficient is

$$B_{\text{eff}} = \frac{1}{2}\int \frac{d\mathbf{u}}{4\pi}\int \frac{d\mathbf{u}'}{4\pi}\int d\mathbf{R}[1-\exp(-W_{\text{eff}}/kT)], \tag{20}$$

where W_{eff} is the effective potential of interaction (potential of mean force) for two rodlike segments with orientational

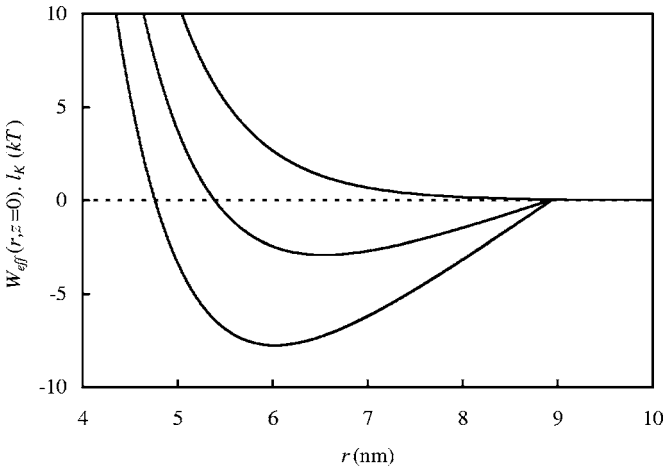

FIG. 6. Effective potential of interaction $W_{\mathrm{eff}}(r,z=0)$ of rodlike, parallel DNA Kuhn segments (length $l_K=100$ nm) at a distance r. Parameter values: $n_s=0.15M$, $a_p=2.5$ nm, and $Z=15$. From top to bottom: $\phi_p=0$, $\phi_p=0.15$, and $\phi_p=0.25$. At the highest protein volume fraction, the effective virial coefficient of the DNA segments, $B_{\mathrm{eff}}\approx 0$.

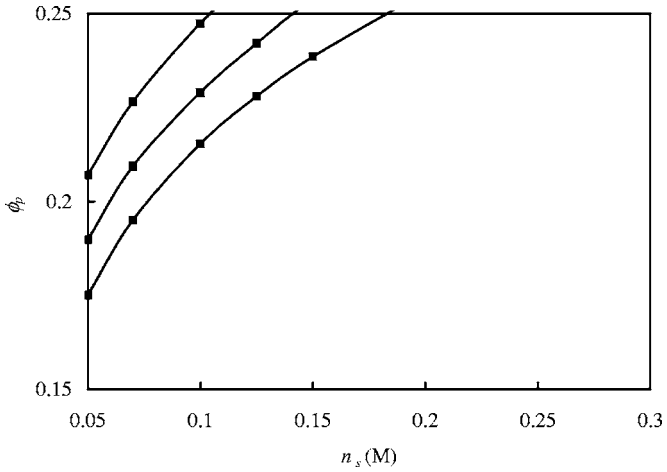

FIG. 7. Instability thresholds for mixtures of globular proteins and linear DNA: critical protein volume fraction ϕ_p as a function of ionic strength n_s. Influence of protein size for proteins far away from the theta point $B_{pp}=0$. The protein surface charge density is kept constant. From top to bottom: $a_p=3.5$ nm and $Z_{\mathrm{eff}}=30$; $a_p=2.5$ nm and $Z_{\mathrm{eff}}=15$; $a_p=2$ nm and $Z_{\mathrm{eff}}=10$. Solid lines are guides to the eye.

unit vectors $\mathbf{u}$ and $\mathbf{u}'$ and centers of mass connected by a vector $\mathbf{R}$. The latter is repulsive at short distances and attractive at larger distances, as illustrated in Fig. 6. To allow for a simple estimate of the instability threshold, we assume that there exists an effective diameter D_{eff} for the rodlike polyelectrolyte segments, such that for center-to-center distances $r<D_{\mathrm{eff}}$, repulsion dominates, whereas for $r>D_{\mathrm{eff}}$, attraction dominates. Next we separate the effective virial coefficient into repulsive and attractive contributions,

$$B_{\mathrm{eff}}=B^R_{\mathrm{eff}}+B^A_{\mathrm{eff}}. \tag{21}$$

The repulsive contribution to the effective virial coefficient is approximated assuming that for $r<D_{\mathrm{eff}}$, the segments behave as long hard cylinders with a diameter D_{eff},

$$B^R_{\mathrm{eff}}\approx\frac{\pi}{4}D_{\mathrm{eff}}l_K^2. \tag{22}$$

For D_{eff} we here use the distance at which the total interaction of parallel rodlike polyelectrolytes, $V(r)$, passes through zero (see Fig. 6). Note that any remaining electrostatic polyelectrolyte-polyelectrolyte repulsion at $r>D_{\mathrm{attr}}$ is completely negligible, as is also clear from Fig. 6.

The attractive contribution to the effective virial coefficient is

$$B^A_{\mathrm{eff}}\approx\frac{1}{2}\int\frac{d\mathbf{u}}{4\pi}\int\frac{d\mathbf{u}'}{4\pi}\int_{D_{\mathrm{eff}}<|\mathbf{R}|<D_{\mathrm{attr}}}d\mathbf{R}[1-\exp(-W_{\mathrm{eff}}/kT)]. \tag{23}$$

When attraction is significant, this integral is dominated by the contributions of the parallel configurations, with angles θ between the segments $\theta<D_{\mathrm{attr}}/l_K$.[25] The effective interaction for these configurations only is a function of the perpendicular and parallel distances r and z between the centers of mass of the (nearly parallel) segments. The total interaction is proportional to the length of the parallel sections,

$$W^{\mathrm{eff}}(r,z)=l_KV(r)(1-|z|/l_K)\quad\text{for parallel segments}. \tag{24}$$

This gives

$$B^A_{\mathrm{eff}}\approx 2\pi^2D^2_{\mathrm{attr}}l_K^{-1}\int_{D_{\mathrm{eff}}}^{D_{\mathrm{attr}}}\times rdr\left\{1-\frac{kT}{V(r)l_K}\left[1-\exp\left(-\frac{V(r)l_K}{kT}\right)\right]\right\}. \tag{25}$$

It should be emphasized that this approximation for the attractive contribution to B_{eff} only works if the attraction is strong indeed, i.e., when the attractive potential of interaction of neighboring parallel rods is many times kT. Our aim is to determine the onset of instability, and this certainly requires a rather strong attraction, such that we expect that the approximation is applicable in our case.

An estimate for the instability threshold is obtained by setting $B_{\mathrm{eff}}\approx 0$. Using this criterion, we have mapped the approximate regions of stability and instability for linear DNA as a function of protein volume fraction and ionic strength.

First consider the effect of protein size, for proteins with a relatively high charge, such that they are far away from their theta point ($B_{pp}=0$). As shown in Fig. 7, with increasing protein size, the threshold increases rapidly to very large protein volume fractions. Only rather small proteins (radius less than 2.5 nm) could possibly condense DNA in the range of biologically relevant parameter values ($n_s>0.1M, \phi_p<0.25$).

For DNA ψ condensation, the ionic strength dependence is due to electrostatic DNA-DNA repulsion[17] hence increasing the ionic strength promotes condensation. Here, the ionic strength dependence is the reverse: the main effect is that screening reduces the electrostatic protein-protein and

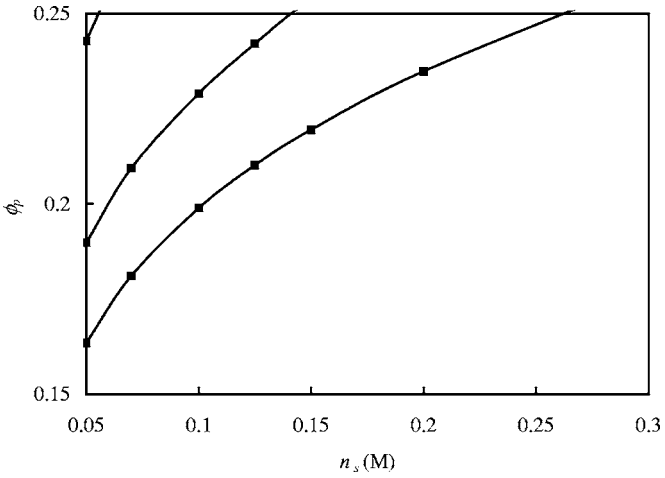

FIG. 8. Instability thresholds for mixtures of globular proteins and linear DNA: critical protein volume fraction ϕ_p as a function of ionic strength n_s. Influence of protein charge, or distance to the theta point, $B_{pp}=0$, at a fixed protein size ($a_p=2.5$ nm). From top to bottom: $Z_{\text{eff}}=10$, $Z_{\text{eff}}=15$, and $Z_{\text{eff}}=20$. The theta point is at $Z_{\text{eff}}=5$. Solid lines are guides to the eye.

protein-DNA repulsions. This reduces the free energy of inserting DNA and reduces the protein-induced depletion attraction between DNA segments.

Even more dramatic than the effect of protein size is the effect of protein charge, or equivalently, protein solubility. Upon decreasing the protein charge, the short-range attraction becomes more and more important until, for some charge, one reaches the protein theta point, $B_{pp}=0$. Below the theta point, protein dimers, trimers, etc., will start to form. As shown in Fig. 8, decreasing the charge (or equivalently, the protein solubility) has a dramatic effect on the instability threshold:long before the theta point has been reached, the threshold disappears from the biologically relevant parameter range of $n_s>0.1M$ and $\phi_p<0.25$. The origin of this behavior is the fact that decreasing the protein solubility also decreases the free energy of inserting DNA in the protein solution, and hence the protein-induced depletion attraction between DNA segments.

DISCUSSION

Whereas Castelnovo and Gelbart[18] find no condensation for $a_p>1.5$ nm, we do find possible instability for globular proteins with small but realistic sizes of approximately 2 nm. A possible explanation for this difference is the fact that while these authors take into account interactions between multiple proteins (second, third and higher order virials), they completely neglect cross interactions between multiple proteins and semiflexible polymer (third and higher order cross virials). The higher order cross interactions significantly increase the insertion free energy, hence the approximation of Castelnovo and Gelbart underestimates the impact of protein depletion. Including not just steric exclusion but also electrostatic repulsions, and nonelectrostatic protein-protein attractions gives further differences.

It should also be kept in mind that the threshold derived from the effective second virial coefficients of DNA segments is actually a lower bound:depletion attraction among multiple DNA segments may give rise to instability when the effective second virial coefficient is still positive.[25,26]

In his calculations for supercoiled DNA, Odijk[20] neglects protein-protein and protein-protein-DNA interactions and argues that this should give a correct leading order estimate, since protein volume fractions inside bacterial cells are only O(0.1). Our detailed calculations suggest that, in fact, protein-protein and protein-protein-DNA interactions have a significant impact on the precise location of the phase boundaries. It would be interesting to see how including these DNAs would shift the computed phase boundaries for supercoiled DNA.

But, under physiological conditions, many proteins are not too far away from their isoelectric points, and often show significant oligomerization. We may hypothesize therefore that on average, globular protein in living cells are close to their theta point ($B_{pp}=0$). Arguments that this is indeed generally the case for prokaryotes were recently put forward by Braun *et al.*[27] Viewed from this perspective, the approximations of Odijk are fully justified, since our Monte Carlo simulations clearly show that for $B_{pp}=0$, the insertion free energy and the depletion attraction are indeed linear in the protein concentration all over the "physiologically" relevant range of $\phi_p<0.25$.

For linear DNA this would imply that globular proteins under "physiologically relevant conditions" (i.e., with on average $B_{pp}=0$) will typically not cause condensation, in agreement with the experimental data of Murphy and Zimmerman[19] who find no condensation for serum albumin. However, small, highly charged proteins may condense linear DNA at relatively low ionic strengths.

Going to more complex systems, the situation is different: as has been emphasized by Murphy and Zimmerman: depletion interactions may not always be strong enough to cause condensation by themselves, but the same may hold for other factors, such as low concentrations of multivalent cations or certain DNA-binding proteins. In combination, these factors may still lead to condensation.

[1] P. Bolhuis and D. Frenkel, J. Chem. Phys. **101**, 9869 (1994).
[2] G. A. Vliegenthart and H. W. N. Lekkerkerker, J. Chem. Phys. **111**, 4153 (1999).
[3] M. Adams, Z. Dogic, S. L. Keller, and S. Fraden, Nature (London) **393**, 349 (1998).
[4] P. van der Schoot, Macromolecules **31**, 4635 (1998).
[5] L. S. Lerman, Proc. Natl. Acad. Sci. U.S.A. **68**, 1886 (1971).
[6] T. Maniatis, J. H. Venable, Jr., and L. S. Lerman, J. Mol. Biol. **84**, 37 (1974).
[7] U. K. Laemmli, Proc. Natl. Acad. Sci. U.S.A. **72**, 4288 (1975).
[8] Yu. M. Evdokimov, T. L. Pyatigorskaya, O. F. Polyvtsev, N. M. Akimenko, V. A. Kadykov, D. Ya. Tsavankin, and Ya. M. Varshavsky, Nucleic Acids Res. **3**, 2353 (1976); M. Evdokimov, T. V. Naesedkina, V. I. Salyanov, N. S. Badaev, Mol. Biol. **30**, 219 (1996).
[9] K. Minagawa, Y. Matsuzawa, K. Yoshikawa, A. R. Khokhlov, and M. Doi, Biopolymers **34**, 555 (1994); K. Yoshikawa and Y. Matsuzawa, Physica D **84**, 220 (1995).
[10] V. V. Vasilevskaya, A. R. Khokhlov, Y. Matsuzawa, and K. Yoshikawa, J. Chem. Phys. **102**, 6595 (1995).
[11] G. Kleideter and E. Nordmeier, Polymer **40**, 4025 (1999).
[12] J. E. B. Ramos, R. de Vries, and J. R. Neto, J. Phys. Chem. B **109**, 23661 (2005).
[13] H. L. Frisch and S. Fesciyan, J. Polym. Sci., Polym. Lett. Ed. **17**, 309 (1979).
[14] C. B. Post and B. H. Zimm, Biopolymers **18**, 1487 (1979); **21**, 2123

(1982).
[15] A. Yu Grosberg and A. V. Zhestkov, J. Biomol. Struct. Dyn. **3**, 859 (1986); A. Yu. Grosberg, Biopolymers **21**, 2413 (1982).
[16] J. Ubbink and T. Odijk, Biophys. J. **68**, 54 (1995); Europhys. Lett. **33**, 353 (1996).
[17] R. de Vries, Biophys. J. **80**, 1186 (2001).
[18] M. Castelnovo and W. M. Gelbart, Macromolecules **37**, 3510 (2004).
[19] L. D. Murphy and S. B. Zimmerman, Biophys. Chem. **57**, 71 (1995).
[20] T. Odijk, Biophys. Chem. **73**, 23 (1998).
[21] P. Prinsen and T. Odijk, J. Chem. Phys. **121**, 6525 (2004).
[22] M. Aubouy, E. Trizac, and L. Boquet, J. Phys. A **36**, 5835 (2003).
[23] D. Frenkel and B. Smit, *Understanding Molecular Simulations* (Academic, San Diego, CA, 1996), Chap. 7, p. 157.
[24] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran 90* (Cambridge University Press, Cambridge, 1996), Vol. 2.
[25] T. Odijk, Macromolecules **27**, 4998 (1994).
[26] R. P. Sear, Phys. Rev. E **58**, 724 (1998).
[27] F. N. Braun, S. Paulsen, R. P. Sear, and P. B. Warren, Phys. Rev. Lett. **94**, 178105 (2005).