Edinburgh Research Explorer

A simple model for DNA bridging proteins and bacterial or human genomes

A simple model for DNA bridging proteins and bacterial or human genomes: bridging-induced attraction and genome compaction.

J. Johnson[1], C. A. Brackley[1], P. R. Cook[2], D. Marenduzzo[1]

[1]SUPA, School of Physics & Astronomy, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JZ, UK

[2]Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford, OX1 3RE, UK

E-mail: `dmarendu@ph.ed.ac.uk`

Abstract. We present computer simulations of the phase behaviour of an ensemble of proteins interacting with a polymer, mimicking non-specific binding to a piece of bacterial DNA or eukaryotic chromatin. The proteins can simultaneously bind to the polymer in two or more places to create protein bridges. Despite the lack of any explicit interaction between the proteins or between DNA segments, our simulations confirm previous results showing that when the protein-polymer interaction is sufficiently strong, the proteins come together to form clusters. Furthermore, a sufficiently large concentration of bridging proteins leads to the compaction of the swollen polymer into a globular phase. Here we characterise both the formation of protein clusters and the polymer collapse as a function of protein concentration, protein-polymer affinity and fibre flexibility.

1. Introduction

Within living cells, either in bacteria or in eukaryotes, DNA rarely exists in its naked form. Instead, it is associated with DNA-binding proteins, which are involved in processes such as transcription and gene regulation, as well as in genome packaging and organisation [1, 2]. The best characterised example of packaging comes from eukaryotes, where the first level of chromosome compaction is provided by the wrapping of DNA around histone octamers in a structure called the nucleosome [3, 4]; strings of nucleosomes then fold into larger structures which form the chromatin fibre [5]. In bacteria, histone-like proteins also determine genome structure; an example is the H-NS protein, which forms dimers [6, 7] that bind non-specifically to AT-rich DNA, and this binding sometimes leads to gene silencing [8]. DNA-associated proteins are also involved in gene regulation; for instance, some eukaryotic proteins compact the fibre into transcriptionally-inactive heterochromatin, whilst others remodel active euchromatin, deeply affecting the local and global folding of the fibre [9, 10].

In general, most DNA-binding proteins are positively charged [1], and this provides a non-specific attractive interaction with negatively-charged DNA. Additional sequence-dependent interactions allow transcription factors and other regulatory proteins to recognise and bind tightly to their DNA targets [11]. The combination of non-specific and specific binding makes it possible for proteins to search for their targets more efficiently via *facilitated diffusion* [11–13]. – a combination of 3-dimensional diffusion in the bulk and 1-dimensional diffusion along the DNA (when the protein is non-specifically bound).

The number of DNA-binding proteins in a cell is extremely large. For instance, in *E. coli*, the bacterial genome is approximately 4.6 Mbp long, and there are ∼30,000 DNA-binding proteins – potentially enough for one protein to bind every ∼ 150 bp [14]. In eukaryotes the numbers are also large, with roughly one nucleosome every 200 bp [1], and tens of thousands of transcription factors [9]. Therefore, it is important to understand the collective (statistical) behaviour of a large number of DNA-binding proteins interacting with a substrate genome, whether bacterial DNA or eukaryotic chromatin. In this work we use a simple bead-and-spring polymer model which, with a suitable choice of parameters, can be used to describe both of these systems. For eukaryotic chromatin we adopt the common model [15–18] in which the nucleosomes arrange into a fibre with diameter 30 nm with the DNA compacted such that there is ∼ 3 kbp per 30 nm length‡; *in vitro* measurements show such fibres to have persistence lengths in the range 40-200 nm [20]. For the case of the bacterial chromosome, it is more appropriate to model naked DNA; the relevant parameters are the hydration thickness of 2.5 nm, and the well characterised persistence length of 50 nm (since the persistence

---

‡ Fibres of 30 nm diameter have been well studied *in vitro*, however the precise structure of such fibres is still a subject of debate [19]. Furthermore, recent work has questioned the existence of the 30-nm fibre *in vivo* for some cell types [10]. Our results, however, will remain qualitatively the same whatever the thickness of the underlying chromatin fibre of which our polymer is a coarse grained representation.
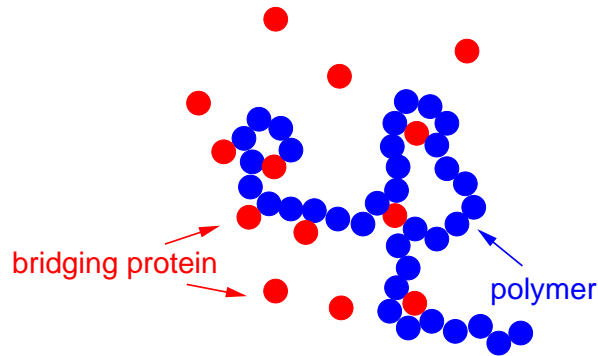
Figure 1. Schematic of the model we consider: DNA or chromatin is modelled as a bead-and-spring polymer (here a flexible polymer is depicted, relevant for euchromatin, see text), while bridging proteins are modelled as spheres which interact with all beads in the polymer with an attractive interaction (see Methods).

length is orders of magnitude longer than the thickness, such a polymer is described as semi-flexible). A schematic diagram is shown in Fig. 1. We model DNA-binding bridging proteins as simple isotropic spheres; like many transcription factors (which can act singly or in conjunction with others), these spheres can simultaneously interact with two or more DNA segments and so form a bridge between the two. Such bridges can then form loops in the intervening DNA or chromatin fibre, and this can play an important part in gene regulation [1,9]. Examples of transcription factors that can act as such bridges to loop DNA include bacterial H-NS [6,7,21], and eukaryotic CTCF [22].

The model considered here is the simplest version of those studied recently in Ref. [15], where – despite the lack of an explicit interaction between one protein and another, or between one DNA segment and another – bound proteins come together to form clusters. Such clustering is driven by a "bridging-induced-attraction" which relies on proteins being able to bind to two or more different DNA segments, and results from a combination of entropic and kinetic effects (the relative balance between which, is slightly different in the different cases considered). This clustering is a generic phenomenon: it occurs with flexible and semi-flexible polymers (representing chromatin or DNA), with both specific and non-specific binding, and with proteins of different size. The model is also similar to the "strings-and-binders" model of Refs. [16,17], with the main differences being that their model is defined on a 3D lattice, they specifically considered chromatin (i.e., the polymer is flexible), and the protein-chromatin interaction was switched on for a only a fraction of chromatin beads (so only specific binding was modelled). Those studies focussed on polymer behaviour and found that increasing protein concentration drives chromatin compaction, and that the exponents reflecting the frequencies of chromatin contacts were (at a suitable protein concentration), similar to those observed in HiC experiments [23,24] (previously also predicted by a polymer-only model known as the fractal globule [18,23]). We also note the similarity between protein induced DNA compaction, and that induced by the

presence of smaller, charged ligands (see e.g. Ref. [25]).

Here, we provide complementary simulations to those described in Refs. [15–17], and systematically study both the clustering dynamics and the bridging-induced compaction of DNA as a function of protein concentration and protein-DNA affinity. The picture emerging from our results is as follows. At low protein concentrations, the polymer is weakly affected, and the most visible phenomenon is a bridging-induced clustering (as in Ref. [15]). At high protein concentrations, the polymer collapses into a globular state (as in Refs. [16, 17]). This bridging-induced compaction is similar to the theta collapse of polymers in a poor solvent [26], and can be quantitatively followed by measuring, for example, the radius of gyration of the polymer. Flexible and semi-flexible polymers behave similarly. We also find that the clusters typically fuse and coarsen, to eventually yield only one cluster at equilibrium. The coarsening dynamics is fast at low and high protein concentrations, but becomes extremely slow at intermediate concentrations; it is also affected by the polymer persistence length (i.e., its stiffness). We believe our simple model can provide a deeper theoretical understanding of the physics underpinning the compaction of bacterial and eukaryotic chromosomes [15–17].

2. Methods

To study the phase behaviour and kinetics of the system, we perform molecular dynamics (MD) simulations via the LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator) software package [27]. We coarse grain the system by treating DNA, or chromatin, as a bead-and-spring polymer. DNA-binding proteins are modelled as spheres. LAMMPS was run in Brownian dynamics (BD) mode, where a molecular dynamics algorithm is used with a stochastic thermostat, which models the thermal fluctuations and viscosity of an implicit solvent. For computational efficiency, hydrodynamic interactions are neglected. This coarse-grained approach – where molecules are represented by only a small number of component spheres – allows our simulations to cover much larger length and time scales than studies which include greater (e.g., atomistic) detail [28–31].

Translational motion of the ith bead (whether a polymer bead or a protein) evolves according to the Langevin equation

$$m_i \frac{d^2\mathbf{r}_i}{dt^2} = \mathbf{F}_i - \gamma_i \frac{d\mathbf{r}_i}{dt} + \sqrt{2k_B T \gamma_i}\,\boldsymbol{\eta}_i(t), \tag{1}$$

where k_B is the Boltzmann constant, T is the temperature, and $\mathbf{r}_i$ is the position of the bead, which has mass m_i; $\mathbf{F}_i$ is the net force on the bead due to interactions with the rest of the system, $\gamma_i \equiv \gamma$ is the friction due to the solvent, and $\boldsymbol{\eta}(t)$ is a vector representing random uncorrelated noise, such that $\langle \eta_\alpha(t) \rangle = 0$, $\langle \eta_\alpha(t)\eta_\beta(t') \rangle = \delta_{\alpha\beta}\delta(t - t')$.

To describe the interactions between the beads we use phenomenological interaction potentials (e.g., truncated Lennard-Jones) since these offer a significant improvement in computational efficiency over more realistic potentials. This choice means that we do not consider any additional electrostatic repulsion between atoms, but since at physiological

salt concentrations the Debye screening length is of order 1 nm, we do not expect any significant effect on our results.

For the polymer chain model, the i-th bead is connected to the $i+1$th with a finitely extensible non-linear elastic (FENE) spring described by the potential

$$U_{\text{FENE}}(r_{i,i+1}) = U_{\text{WCA}}(r_{i,i+1}) - \frac{K_{\text{FENE}}R_0^2}{2} \log\left[1 - \left(\frac{r_{i,i+1}}{R_0}\right)^2\right],\tag{2}$$

where $r_{i,i+1} = |\mathbf{r}_i - \mathbf{r}_{i+1}|$ is the separation of centres of the beads. Here the first term represents a hard steric interaction with prevents adjacent beads from overlapping (the Weeks-Chandler-Andersen (WCA) potential – see Eq. (3) below), and the second gives the maximum extension of the bond R_0. Throughout we use $R_0 = 1.6\ \sigma$, and set the bond energy $K_{\text{FENE}} = 30\ k_BT/\sigma^2$, where σ is the polymer bead diameter (see below for the values we use in the different models).

Steric interactions between non-adjacent DNA beads, and between any two proteins, are also given by the WCA potential

$$U_{\text{WCA}}(r_{ij}) = \begin{cases} 4k_BT\left[\left(\frac{d_{ij}}{r_{ij}}\right)^{12} - \left(\frac{d_{ij}}{r_{ij}}\right)^6\right] + k_BT & r_{ij} < 2^{1/6}d_{ij}, \\ 0 & \text{otherwise.} \end{cases}\tag{3}$$

Here d_{ij} is the mean of the diameters of the two spheres, i.e., for two polymer beads $d_{ij} = \sigma$.

The bending rigidity of the polymer is introduced via a Kratky-Porod potential for every three adjacent DNA beads

$$U_{\text{BEND}}(\theta_B) = K_{\text{BEND}}\left[1 - \cos(\theta_B)\right],\tag{4}$$

where θ_B is the angle between three consecutive beads (measured such that $\theta_B = 0$ if the three beads are collinear), and K_{BEND} is the bending energy. The persistence length in units of σ is given by $l_p = K_{\text{BEND}}/k_BT$ (see below for the values we use in the different models).

The interaction between protein bridges and DNA/chromatin is defined as a shifted truncated Lennard-Jones potential. To define this, we start from the usual Lennard-Jones potential,

$$U_{\text{LJ}}(r_{ij}) = 4\epsilon'\left[\left(\frac{d_{ij}}{r_{ij}}\right)^{12} - \left(\frac{d_{ij}}{r_{ij}}\right)^6\right].\tag{5}$$

If we truncate this potential at $d_{ij} = r_{\text{thr}}$ and require it to be continuous, we obtain

$$U_{\text{trLJ}}(r_{ij}) = \begin{cases} U_{\text{LJ}}(r_{ij}) - U_{\text{LJ}}(r_{\text{thr}}) & r_{ij} < r_{\text{thr}}, \\ 0 & \text{otherwise.} \end{cases}\tag{6}$$

Equation 6 reduces to the purely repulsive WCA potential for $r_{\text{thr}} = 2^{1/6}d_{ij}$, while if $r_{\text{thr}} > 2^{1/6}d_{ij}$ an attractive tail sets in. The parameter ϵ' controls the strength of the interaction, but due to the shifting of the function this is not the same as the energy minimum of the potential; for clarity we use the unprimed symbol ϵ to denote the true

value of the energy at the minimum. For simplicity, we take the size of the protein equal to that of the DNA/chromatin bead (considering different sizes would be more appropriate for most DNA-binding proteins, but this does not qualitatively affect the results we present here).

As described above, to model naked DNA we use $\sigma = 2.5$ nm (one bead is therefore ~ 7.35 bp) and set $l_p = 20\ \sigma = 50$ nm. For chromatin, we use $\sigma = 30$ nm, taking this to contain 3 kbp of packaged DNA, and use a persistence length of 90 nm which is appropriate for transcriptionally active, open euchromatin [15].

The simulation time units are defined in terms of the friction γ (see Eq. 1; we use $\gamma = 2$ unless otherwise stated), which is in turn related to the diffusion constant of a bead of size σ via $D = k_B T/\gamma$. A natural time scale is then the Brownian time, τ_B, which is the order of magnitude of the time needed for a bead to diffuse across its own size; for a bead of diameter σ, $\tau_B = \sigma^2/D$. One Brownian time in simulation units therefore corresponds to a different time-scale for DNA and chromatin, ~ 36 ns and ~ 0.6 ms respectively – estimated using a viscosity for the cytosol in bacteria or the nucleoplasm for chromatin, of 1 cP (DNA, semi-flexible polymer), or 10 cP (chromatin, flexible polymer) respectively. In the results below all times given are mapped to physical units.

3. Results

We begin the presentation of our results with the case of a flexible polymer. As detailed above, this is relevant for active euchromatin [15,32], (we assume a persistence length $l_P = 90$ nm, chosen towards the lower end of the 40-200 nm range determined experimentally [20]). The key parameters in our system are: protein (bridge) concentration, c_b, polymer concentration, c_p, and protein-polymer affinity, ϵ, which measures the minimum of the protein-DNA interaction potential §. Instead of changing c_p and c_b independently, we fix c_p and vary $x \equiv c_p/c_b$; as we shall see, varying this ratio leads to a change in behaviour from the bridging-induced attraction of Ref. [15] to the binder-induced compaction of Refs. [16,17]. Note that Refs. [16,17] used comparatively small polymers, where only every second or third monomer interacts with the proteins; under these conditions a small x corresponds to only single digit numbers of proteins – hence in that work they mostly consider $x > 1$ which is larger than the values we consider here.

Figures 2A-C show snapshots taken from a simulation where a 5000-bead polymer (corresponding to a 15-Mbp euchromatin fibre, see Methods) interacts with 1000 proteins (hence $x = 0.2$). For the relatively large value for the affinity between protein

---

§ Another parameter is the ratio between the polymer size, quantified by its radius of gyration R_g, and the size of the (cubic) simulation box, L. If $R_g \ll L$, this corresponds to the dilute regime, an R_g of the same order of magnitude as L is semi-dilute, whereas higher R_g (say for a c_p corresponding to a volume fraction of 10% or more) corresponds to concentrated polymer. We work in the dilute and semi-dilute regimes, and did not find any qualitative differences between their behaviours.

and DNA used here ($\epsilon = 2.83\ k_B T$), the bridges stick essentially irreversibly to the chromatin. Clusters of proteins form early on during the simulation (Fig. 2A) due to the bridging-induced attraction reported in Ref. [15]; later, these coarsen (Fig. 2B,C) to leave eventually only one aggregate in steady state (the number of clusters versus time is shown in Fig. 2D). We define two proteins as belonging to the same cluster if their centre-centre separation is below a threshold; for the case of chromatin this is chosen to be 90 nm (3 σ; Figs. 2-4). The full dynamics corresponding to Fig. 2 are shown in Suppl. Movie 1, and this reveals that the coarsening occurs predominantly by fusion of clusters which meet stochastically. There is little evidence of Ostwald ripening, which would lead to the growth of a cluster at the expense of another one which would shrink – presumably because the interaction between bridges is large enough that the disintegration of an aggregate, once formed, is very unlikely. The final clustered state is also robust to different initial conditions, e.g., it still appears if the proteins are initially bound at regularly spaced positions along a swollen polymer – see Ref. [15].

Our simulations show that the fraction of proteins involved in the clustering is very high even for the relatively low protein concentration considered in Fig. 2: this means that once proteins form bridges, they come together through the bridging-induced attraction (i.e., there are very few isolated bridges – see also Fig. 3). Due to the coarsening of the initial clusters into a large single final aggregate, the average size of the cluster is very close to that of the total number of bridges (Fig. 2D). We have studied here protein concentrations in the range $0.1 \leq x \leq 0.5$; while we expect coarsening until a single aggregate is left for all x that we studied, for intermediate x our simulation time (10^6 Brownian times) is not sufficient to observe the completion of this process, so that the state with multiple clusters could be dynamically stabilised$\|$. As noted above the coarsening occurs through fusion of clusters, which must come together stochastically via diffusion; for small clusters, their diffusion is relatively fast, but as more proteins are added, the size of the initial clusters grows, and the kinetics of coarsening become very slow. For much larger x, the proteins initially bind all over the chromatin: there are more initial aggregates which are closer together, so again the coarsening is faster.

Although for the affinity considered in Fig. 2 proteins form clusters whatever the value of x, the impact of protein binding on genomic structure is modest for low x, but is massive for large x. This can be appreciated, for instance, by looking at the plot of the radius of gyration of the polymer, R_g, versus x (Fig. 3), which shows that increasing x induces a transition from an overall open, swollen phase with large R_g, to a compact one, with small R_g. Although we do not measure here the exponents associated with this transition (which would require simulations at multiple values of the polymer contour length), we expect that the exponent ν (where $R_g \sim N^\nu$, with N the number of polymer beads) should be close to 0.588 (the value for a swollen polymer [26]), whereas for large x, ν should approach 1/3. For the affinity considered in Fig. 3, the equilibrium state is a coexistence of an open region of the polymer with a globular regions corresponding to

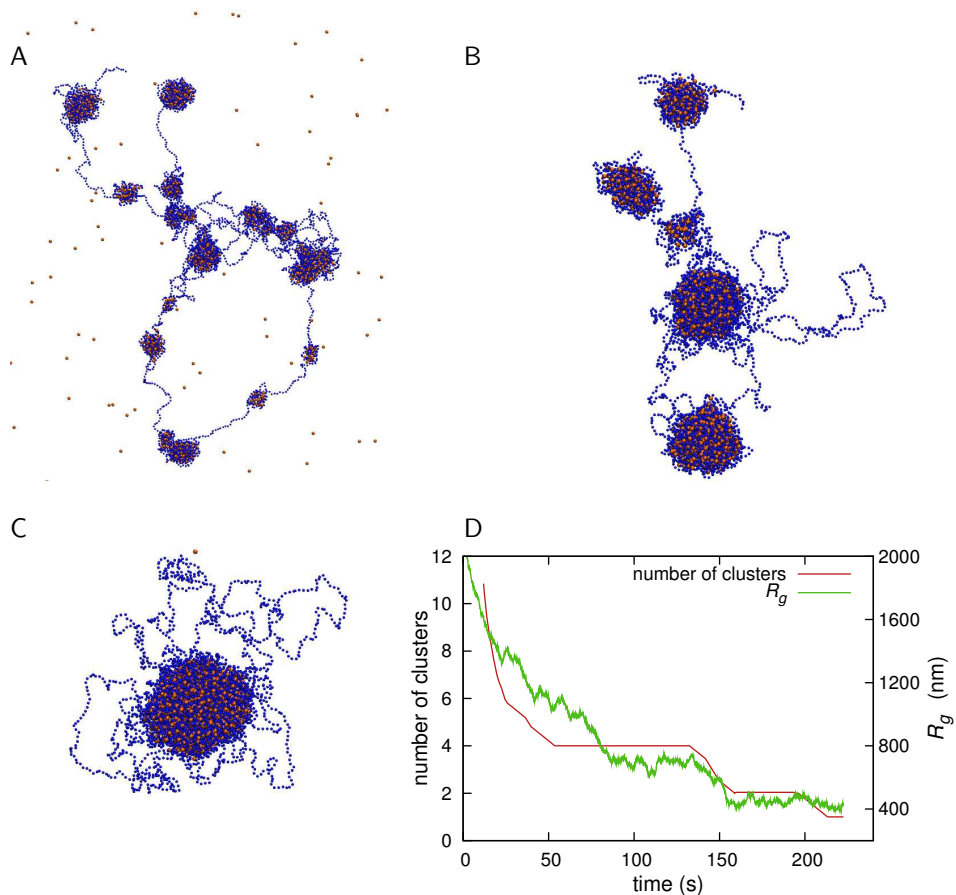$\|$ The coarsening dynamics also depends on the value of r_{thr}.

Figure 2. Panels A-C show snapshots at increasing times during a simulation where 1000 proteins interact with a chromatin fibre represented by 5000 polymer beads (bead diameter $\sigma = 30$ nm, persistence length $l_p = 90$ nm, and x=0.2). Initial conditions were an equilibrium coil configuration for the chromatin, with proteins distributed randomly through a cube of side $L = 500$ nm; at $t = 0$ the protein-polymer attraction was switched on and protein clusters quickly form. Panel D shows how the number of clusters and radius of gyration decrease with time. Proteins interact with the polymer according to Eq. (6) with $\epsilon = 2.83 \ k_B T$ and $r_{\mathrm{thr}} = 60.6$ nm (2.02 σ).

the cluster formed by the bridging-induced attraction. The size of the globule depends on x, i.e. increasing x increases the globule size until it eventually takes over the whole fibre.

The bridging-induced compaction observed in Fig. 3 has previously been documented in Refs. [16, 17]; in some of the cases considered in Ref. [16] the collapsed state could also be reached by changing ϵ at large x. In this case, the switch between open and collapsed phase appears similar to that of the theta collapse of a polymer in a poor solvent, which can be triggered, for instance, by an increase in the effective attractive interaction between the monomers in the polymer [26].

The reorganisation of the polymer which occurs as a result of the bridging-induced attraction provides a useful framework for understanding several observations in chromosome biology. Recent HiC experiments have uncovered a power law governing the
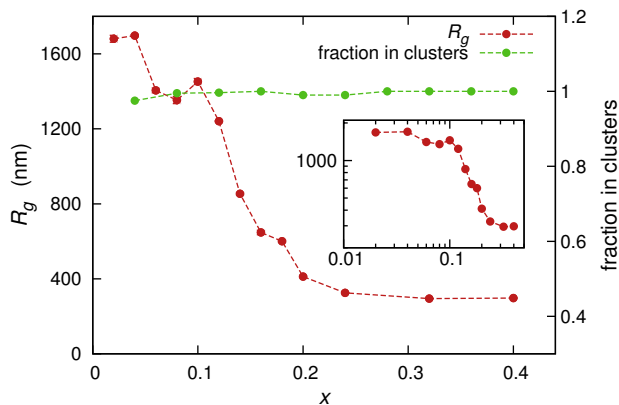
Figure 3. Plot showing the radius of gyration of the chromatin fibre at the end of a simulation for different values of x (after a run time of at least 2.5 s, and then averaged over at least a further 25 ms; error bars show the standard error in this mean). The inset shows a log-log plot for R_g vs. x, which can be compared e.g. to Fig. 1D in Ref. [17]. Also shown is the proportion of proteins found in clusters; note that even before the compaction sets in, most bridges are clustering. The length of the polymer is fixed at 5000 beads, and the other parameters are as in Fig. 2.
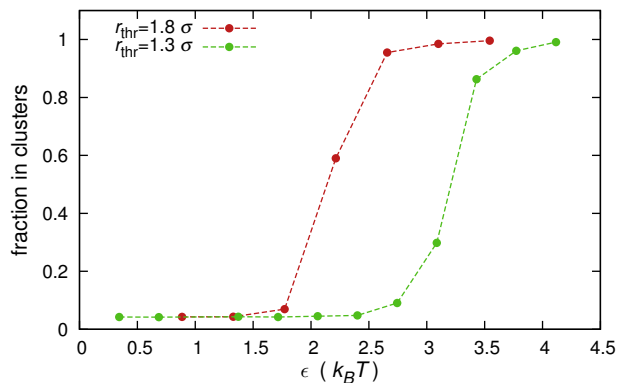

Figure 4. Plot of the fraction of protein bridges within clusters as a function of affinity, for a flexible polymer, and for two values of the range, r_{thr} (see Methods). Other parameters are: $c_p = 0.26\%$ (in volume fraction), and $x = c_b/c_p = 0.08$.

decay in contact probability with distance along the polymer [23, 24], with an exponent which depends on cell type; Ref. [17] showed that by changing the protein concentration, c_p, it was possible to vary the power law dictating the decay of contacting probability within a simulated polymer. Also, it is known that chromatin fibres are disordered, with compact heterochromatic regions interspersed amongst open euchromatic ones [10]; the coexistence of a cluster or globule state with a more open, or swollen region reported here provides a generic pathway to drive segregation of different chromatin states.

Figure 4 shows the effect that different protein-DNA affinities have on the fraction of proteins in clusters, for a small value of x. This fraction is close to zero for low affinity, and increases sharply above a critical value, $\epsilon > \epsilon_c$ (where $\epsilon_c \sim 2.2\ k_BT$ for
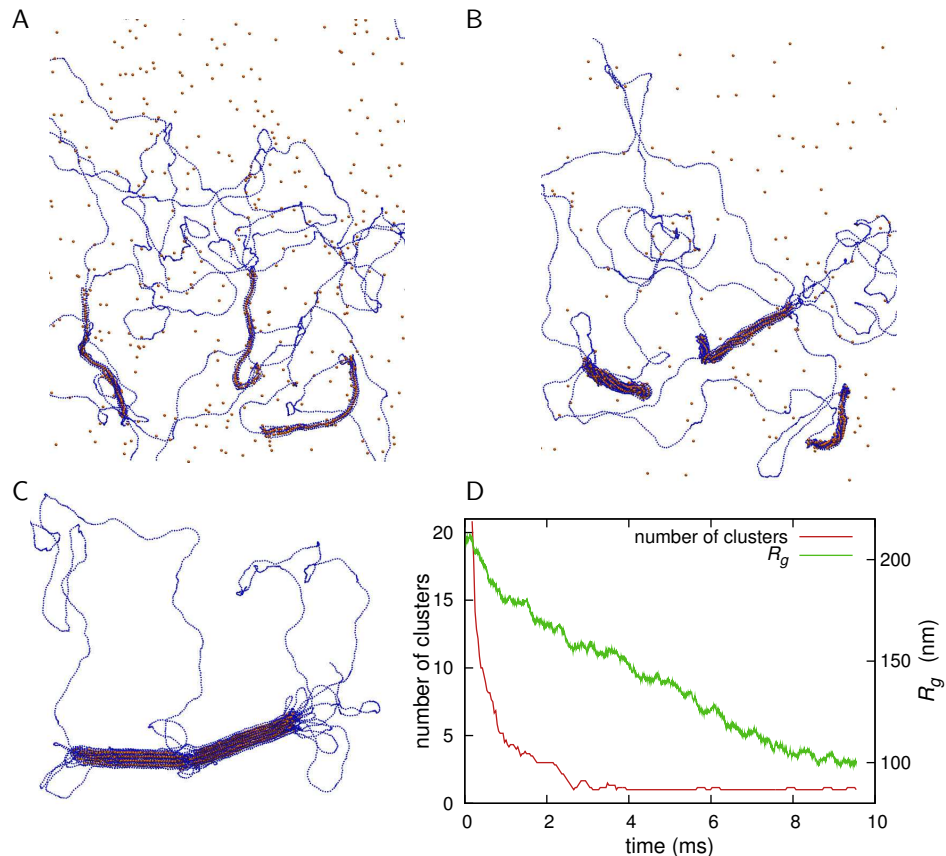
Figure 5. Similar plots to those in Fig. 2, but for the case of a semi-flexible polymer representing naked DNA (bead diameter $\sigma = 2.5$ nm, persistence length $l_p = 50$ nm). Panels A-C show snapshots of the clustering at different times for 800 proteins interacting with a 5000 bead (36.8 kbp) DNA molecule ($x = 0.16$). Panel D shows how the number of cluster and the radius of gyration change with time for the same similar simulation. Proteins interact with the DNA according to Eq. (6) with $\epsilon = 2.83\ k_B T$ and $r_{\text{thr}} = 5.05$ nm (2.02 σ).

$r_{\text{thr}} = 54$ nm (1.8 σ), and $\epsilon_c \sim 3.1\ k_B T$ for $r_{\text{thr}} = 39$ nm (1.3 σ); we expect in general ϵ_c to depend on both c_p and c_b separately). Consistent with the results of Fig. 3, we only observe a modest (35-40%) decrease in polymeric size around ϵ_c (this is because the value of x is below that required for full compaction).

We now discuss the case where the polymer is semi-flexible – modelling proteins like H-NS binding non-specifically to bacterial DNA [6,7]. We find that the polymer rigidity plays an important role. First, Figs. 5A-C (and see also Suppl. Movie 2) show that the clusters which arise due to the bridging-induced attraction now are cylindrical in shape, rather than spherical. For the semi-flexible case we define the clustering threshold as 3.5 nm (1.4 σ; Figs. 5-7). We still observe coarsening, but at a lower rate compared to the flexible polymer (see Fig. 5D ¶); our simulations are also inconclusive as to whether

¶ Note that the time-scales in Fig. 2D and Fig. 5D are different due to the different sizes of DNA and chromatin, however the total simulated time in these two figures correspond to the same number
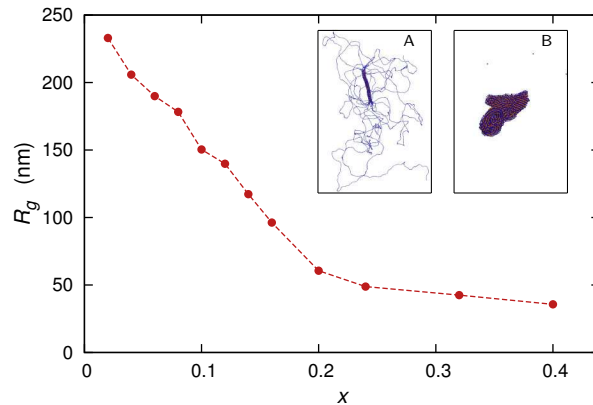
Figure 6. Plot showing the radius of gyration of the DNA at the end of a simulation for different values of x (after a run time of at least 3 ms, and then averaged over at least a further 0.03 ms; error bars show the standard error in this mean). The length of the DNA is fixed at 5000 beads, and the other parameters are as in Fig. 5. The insets show snapshots corresponding to $x = 0.04$ (A) and $x = 0.4$ (B).

a single cluster or several remain in equilibrium. The row-like clusters are qualitatively similar to the structures which have been inferred for H-NS interacting with DNA on the basis of atomic force microscopy [7]. As for the flexible case in Fig. 3, increasing x eventually leads to bridging-induced compaction of the DNA (Fig. 6). However, unlike in the flexible case x decreases only gradually (rather than a sharp fall at a critical value as in Fig. 3); possibly, this is because in the semi-flexible case the bridges are more likely to cause long-distance looping, which can more efficiently compact the polymer than the local folding which is more common in the flexible fibre.

Finally, the effect of changing protein-DNA affinity is similar to the flexible case, and the critical threshold beyond which clustering sets in, for the same concentration as in Fig. 4, is now slightly higher at $\epsilon_c \sim 3.4\ k_B T$ (for $r_{\mathrm{thr}} = 3.25$ nm (1.3 σ); compare Fig. 7 and Fig. 4).

4. Discussion and conclusions

To summarise, here we have studied a simple statistical physics model for a DNA molecule or chromatin fibre interacting with a solution of DNA-binding proteins which can bind to the DNA at more than one point to form bridges. We varied the flexibility of the polymer, as well as the concentration and the polymer binding affinity of the proteins.

For a sufficiently attractive interaction between DNA and the bridges, we observe protein clustering (see e.g. Fig. 2 and Fig. 5); the shapes of the resulting aggregates are spherical (for flexible polymers) or elongated (for stiffer ones). This clustering is reminiscent of the aggregation seen experimentally in mixtures of DNA and

of Brownian times, $\tau_B = \sigma^2 \gamma / (k_B T)$. We note also that in reality the times-scales will increase with polymer length (the exact scaling will depend on hydrodynamic interactions and polymer confinement).
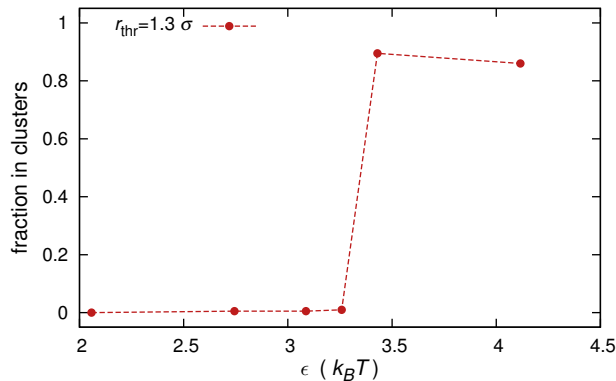
Figure 7. Plot of the fraction of bridges within clusters as a function of affinity, for a semi-flexible polymer, and an interaction range, or threshold, $r_{\mathrm{thr}} = 3.25$ nm ($1.3\ \sigma$; see Methods). Other parameters are as in Fig. 4.

nanoparticles (representing 'synthetic proteins', [33]), and is an example of the bridging-induced attraction discussed in Ref. [15]: it arises because bridging enhances the local DNA density, which in turn recruits more bridges – setting up a positive feedback loop. The formation of clusters may be described as a phase transition, which is triggered by an increase in protein-DNA affinity (Fig. 4 and Fig. 7). Simulations suggest that separate clusters are unstable, as they fuse to give larger aggregates when they meet stochastically. For intermediate values of the protein concentration, such coarsening does not always proceed to leave a single cluster within the duration of our simulations. Presumably, this is due to a kinetic reason: if clusters form in far-away regions of the DNA, then fusion between two of them requires their meeting by chance through diffusion – this is a slow process for large clusters.

The clustering which we generically observe may provide a simple framework to understand genome organisation. For instance, binding to, and forming bridges within, a flexible polymer can yield zones of compact (and transcriptionally inactive) heterochromatin [9, 10] and zones of more open euchromatin, like those found in mammalian chromosomes. Similarly, binding to semi-flexible polymers yields structures reminiscent of those seen experimentally with bacterial DNA and H-NS [6, 7].

In this work we also showed that the concentration of protein bridges plays an important role in the physics of the system. For sufficiently strong protein-DNA interactions, proteins are preferentially bound to the DNA, hence the relevant physical parameter is the ratio between the protein and polymer concentration, or equivalently the linear packing fraction of proteins along the DNA, which we call x. If x is low, bridges cluster and condense the DNA locally, and one is left with a globular region coexisting with swollen regions. When x increases, the swollen region shrinks and the size of the polymer drops sharply. When x is large enough, the whole polymer becomes compacted. These results provide a link between the phenomena described in Ref. [15] (clustering of bridges through the bridging-induced attraction for relatively low values of x), and the findings of Ref. [17] (where x is typically larger than 1, and a switch

between an open and a bridging-induced compacted phase was observed upon variation of either protein concentration or protein-DNA affinity). Ref. [17] also showed that the compaction is linked to a qualitative change in the statistics of contact formation, which may be relevant to the interpretation of HiC experiments on inter-chromosomal contacts [17].

Both the bridging-induced attraction and compaction are likely to be generic phenomena that can be observed with DNA-binding bridges which bind either non-specifically or specifically. It would be of interest to repeat the study provided here both with protein bridges with limited valencies and mixtures of specific and non-specific interactions, and to see how this changes the physics of the bridging-induced compaction. One might expect, for example, that a more specific interaction with DNA or chromatin might lead to stabilization of multiple clusters, segregation of polymer regions with different binding properties, or the formation of local domains of increased polymer interaction — all of which have far reaching implications for genome organisation.

References

[1] B. Alberts *et al.*, *Molecular Biology of the Cell*, Garland Science, New York (2002).
[2] C. R. Calladine, H. R. Drew, *Understanding DNA*, Academic Press Limited, Cambridge (1992).
[3] K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent and T. J. Richmond, *Nature* **389**, 251–260 (1997).
[4] C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder and T. J. Richmond *J. Mol. Biol.* **319**, 1097–1113 (2002).
[5] A. P. Wolffe, *Chromatin Structure and Function*, Academic Press, London (2000).
[6] J. Vreede and R. T. Dame *Biophys. J.* **103**, 89–98 (2012).
[7] R. T. Dame, M. C. Noom and G. J. L. Wuite *Nature* **444**, 387–390 (2006).
[8] B. Lang et al., *Nucl. Acids Res.* **35**, 6330–6337 (2007).
[9] P. R. Cook, *Principles of Nuclear Structure and Function*, Oxford University Press, New York (2001).
[10] Y. Nishino *et al.*, *EMBO J.* **31**, 1644 (2012).
[11] M. Sheinman, O. Benichou, Y. Kafri, R. Voituriez, *Rep. Progr. Phys.* **75**, 026601 (2012).
[12] O. G. Berg, R. B. Winter, and P. H. von Hippel *Biochemistry* **20**, 6929 (1981)
[13] C. A. Brackley, M. E. Cates, D. Marenduzzo, *Phys. Rev. Lett.* **109**, 168103 (2012); *Phys. Rev. Lett.* **111**, 108101 (2013).
[14] G.-W. Li, O. G. Berg, J. Elf, *Nat. Phys.* **5**, 294 (2009).
[15] C. A. Brackley, A. Papontonis, S. Taylor, P. R. Cook, D. Marenduzzo, *Proc. Natl. Acad. Sci. USA* **110**, E3605 (2013).
[16] M. Nicodemi, A. Prisco, *Biophys. J.* **96**, 2168 (2009).
[17] M. Barbieri *et al..*, *Proc. Natl. Acad. Sci. USA* **109**, 16173 (2012).
[18] L. A. Mirny *Chromosome Res* **19**, 37–51 (2011).
[19] G. Li and D. Reinberg *Curr. Opin. Gen. Dev.* **21**, 175 (2011).
[20] J. Langowski, *Eur. Phys. J. E* **19**, 241 (2006).
[21] M. C. Noom, W. W. Navarre, T. Oshima, G. J. L. Wuite and R. T. Dame *Current Biology* **17**, R913–R914 (2007).
[22] J. E. Phillips and V. G. Corces *Cell* **137**, 1194–1211 (2009).
[23] E. Lieberman-Aiden et al., *Science* **326**, 289–293 (2009).
[24] J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu and B. Ren *Nature* **485**, 376–380 (2012).

[25] V. B. Teif, *Biophysical J.* **89**, 25742587 (2005).

[26] P. G. de Gennes, *Scaling concepts in polymer physics*, Cornell University Press, New York (1979).

[27] S. Plimpton, *J. Comput. Phys.* **117**, 1–19 (1995).

[28] T. C. Bishop, (2005) *J. Biomol. Struct. Dyn.* **22**, 673–685.

[29] Ettig, R., Kepper, N., Stehr, R., Wedemann, G., and Rippe, K. *Biophysical Journal,* **101**, 1999–2008 (2011).

[30] Fathizadeh, A., Berdy Besya, A., Reza Ejtehadi, M., and Schiessel, H. *Eur. Phys. J. E* **36(3)**, 1–10 (2013).

[31] Fan, Y., Korolev, N., Lyubartsev, A. P., and Nordenskild, L. *PLoS ONE* **8(2)**, e54228 (2013).

[32] P. R. Cook, D. Marenduzzo, *J. Cell Biol.* **186**, 825 (2009).

[33] A. A. Zinchenko, K. Yoshikawa, and D. Baigl *Phys. Rev. Lett.* **95**, 228101 (2005).