Nonspecific bridging-induced attraction drives clustering of DNA-binding proteins and genome organization

Chris A. Brackley[a], Stephen Taylor[b], Argyris Papantonis[c,d], Peter R. Cook[c,1], and Davide Marenduzzo[a,1]

[a]Scottish Universities Physics Alliance (SUPA), School of Physics, University of Edinburgh, Edinburgh EH9 3JZ, United Kingdom; [b]Computational Research Biology Group and [c]Sir William Dunn School of Pathology, University of Oxford, Oxford OX1 3RE, United Kingdom; and [d]Center for Molecular Medicine, University of Cologne, 50931 Cologne, Germany

Edited by Charles S. Peskin, New York University, Manhattan, NY, and approved August 5, 2013 (received for review February 19, 2013)

Molecular dynamics simulations are used to model proteins that diffuse to DNA, bind, and dissociate; in the absence of any explicit interaction between proteins, or between templates, binding spontaneously induces local DNA compaction and protein aggregation. Small bivalent proteins form into rows [as on binding of the bacterial histone-like nucleoid-structuring protein (H-NS)], large proteins into quasi-spherical aggregates (as on nanoparticle binding), and cylinders with eight binding sites (representing octameric nucleosomal cores) into irregularly folded clusters (like those seen in nucleosomal strings). Binding of RNA polymerase II and a transcription factor (NFκB) to the appropriate sites on four human chromosomes generates protein clusters analogous to transcription factories, multiscale loops, and intrachromosomal contacts that mimic those found in vivo. We suggest that this emergent behavior of clustering is driven by an entropic bridging-induced attraction that minimizes bending and looping penalties in the template.

polymer physics | Brownian dynamics | chromatin looping | nucleosome

DNA in living cells associates with proteins that continuously bind and dissociate. Some proteins affect local structure (such as histones and histone-like proteins), whereas others act globally to compact whole chromosomal segments [such as CCCTC-binding factor (CTCF)] (1–3). Bound proteins also cluster into supramolecular structures; for example, different transcription factors often bind to the same hot spots in the fly genome (4), and active molecules of RNA polymerase II coassociate in transcription factories (5, 6). In the latter case, clustering generates high local concentrations that facilitate production of the appropriate transcripts, as well as organizing the genome in 3D space.

Against this background, biophysicists have begun to model DNA folding driven by DNA-binding proteins (3, 7–12). Usually, the effects of DNA binding are incorporated into an effective potential that influences DNA dynamics; for instance, by stipulating that selected protein-binding regions in the polymer attract each other (11, 12). Here, we use molecular dynamics (MD) to model proteins that diffuse to DNA, bind, and dissociate. In the absence of any explicit mutual attraction between proteins or between monomers in the polymer, we uncover an emergent property of the system: binding spontaneously induces protein clustering and genome compaction. For example, simulations yield structures seen experimentally when proteins representing bacterial histone-like nucleoid-structuring protein (H-NS) (1, 2, 13), gold nanoparticles (14, 15), and nucleosome cores bind to DNA. Using data derived from ChIP coupled to high-throughput sequencing (ChIP-seq) (16), we also model binding of RNA polymerase II and its transcription factor, NFκB, to the appropriate (cognate) sites on four human chromosomes; the two proteins spontaneously cluster into factories that are surrounded by loops that reflect those detected in cells using chromatin interaction analysis with paired-end tag sequencing (ChIA-PET) (16, 17).

Results

Clustering of Bridging Proteins on Binding to Naked DNA. We first consider a solution of spherical DNA-binding proteins (concentration, 0.02% in volume or 42.5 μM and therefore within the range found in vivo) that bind nonspecifically to naked DNA (36.7 kbp), modeled as a semiflexible string of spherical monomers (persistence length, 50 nm) (18) confined within a cube (250 × 250 × 250 nm). Both proteins and monomers have diameters of 2.5 nm, each DNA monomer represents ∼7.35 bp, and no two components can occupy the same volume. To avoid edge effects, we use periodic boundary conditions: if a monomer or protein exits through one face of the cube, it reenters through the opposite one. We assume that each protein is attracted to each DNA monomer if any part of the DNA lies within a shell extending 0.75 nm away from the protein surface. An attractive energy of 4.1 k_BT is large enough to ensure the equilibrium favors the bound complex (a precise determination of the dissociation constant, K_d, which is <0.1 μM, depends critically on the instantaneous DNA conformation modeled). As no interaction beyond steric repulsion is introduced (so no protein is directly attracted to another, and no one monomer in the polymer to another), one might expect proteins to bind homogeneously and diffusely along the DNA (binding of all proteins would only occupy 8% of the contour length). Surprisingly, proteins quickly find each other to cluster into rows, locally

Significance

We use molecular dynamics to simulate reversible binding of proteins to DNA and uncover an unexpected force driving DNA compaction and protein aggregation. In the absence of any explicit interactions between proteins, or between templates, we find proteins aggregate spontaneously to locally organize the genome. The simulations reproduce the structures seen experimentally when small bivalent proteins assemble into rows (like bacterial H-NS protein), larger proteins with eight binding sites into irregular strings (like octameric nucleosomal cores in chromatin fibers), and still-larger complexes representing RNA polymerase II and a transcription factor (NFκB) into clusters surrounded by loops (like transcription factories). We suggest clustering is driven by an entropic bridging-induced attraction that minimizes bending and looping penalties in the template.

Author contributions: P.R.C. and D.M. designed research; C.A.B., S.T., and A.P. performed research; C.A.B., S.T., A.P., P.R.C., and D.M. analyzed data; and P.R.C. and D.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. E-mail: peter.cook@path.ox.ac.uk or dmarendu@ph.ed.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1302950110/-/DCSupplemental.

Downloaded by guest on February 24, 2022

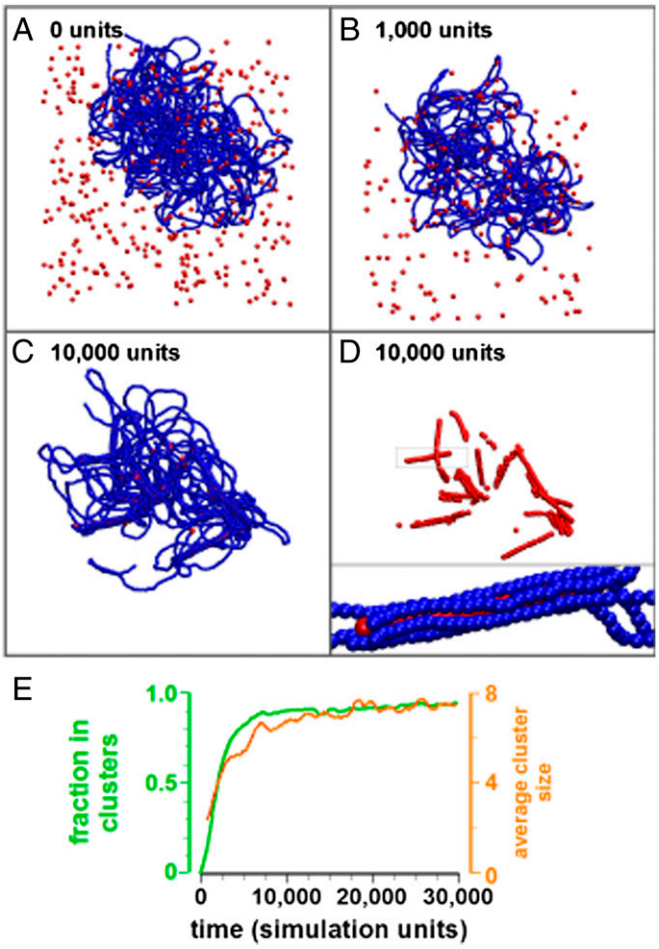

Fig. 1. Small proteins bind to DNA and form rows. MD simulations involving one string of blue beads (2.5 nm diameter) representing 36.75 kbp of DNA (persistence length, 50 nm; volume fraction, 0.26%; the radius of gyration of the unconfined polymer is ~456 nm) (18) interacting with 400 DNA-binding proteins (2.5-nm-diameter red spheres; volume fraction, 0.02%) in a cube (250 × 250 × 250 nm). The interaction energy and range (protein bead center to DNA bead center) were 4.12 k_BT and 3.25 nm, respectively; times shown in all figures are in simulation units (here equivalent to 70 ns/unit, assuming a viscosity of 1 cP). (*A–D*) Snapshots taken at different times. In *D*, only proteins are shown (*Inset*: magnified region with proteins and DNA). As proteins bind, they form into rows, locally folding DNA. (*E*) Both the fraction of beads in clusters and average cluster size increase with time (two bound proteins are in a cluster if center-to-center distance is <3.5 nm).

compacting the polymer (Fig. 1; Movies S1 and S2); an organization like that seen when H-NS binds to bacterial DNA in vitro (1, 13) and perhaps in vivo (2). These clusters grow to reach a steady-state size (Fig. 1*E*), and, once formed, they persist as proteins detach and reattach. The fraction of proteins in clusters (which reaches >90% in the steady state) correlates with the decrease in pairing energy (Fig. S1), indicative of one protein binding to at least two DNA segments.

Results are generic, robust, and independent of initial conditions; similar patterns are observed with interaction energies of 3–10 k_BT, protein concentrations between 1 and 100 μM, and when proteins are prebound randomly to DNA. Clustering is not driven by an effect analogous to the entropic depletion attraction (19) or confinement, because the DNA concentration is so low (i.e., monomer volume fraction, 0.26%). In the absence of proteins, the DNA has a gyration radius of ~456 nm (estimated using a worm-like chain approximation) and therefore is in a semidilute regime; in the presence of proteins, the gyration radius is smaller than the system size. In any case, crowding and confinement hinder clustering, with only ~85% proteins being in clusters at the end of an identical simulation using a higher DNA concentration reflecting that in vivo (i.e., monomer volume fraction, 9.7%; Fig. S2). Moreover, clustering depends on a protein: DNA attraction, because only ~1% of proteins cluster in its absence (Fig. S3).

Because proteins like H-NS have only two DNA-binding sites (1, 13), and as one of our proteins can bind to more than two DNA segments, we repeated the simulations using bivalent proteins that now contain a small DNA-binding region at each pole (Fig. S4*A*); rows again form, although clustering is reduced (Fig. S5). Almost no clustering occurs in simulations with only one binding site per protein (e.g., Fig. S5*E* and Fig. S6 where in the latter case a cluster is defined as two or more 7.5-nm proteins lying within 9 nm of each other, <10% are found in clusters; *Discussion*).

Effect of Protein Size. We next consider fivefold larger proteins that interact with DNA (1.24 k_BT attraction; dissociation constant, <30 nM). There is again no explicit interaction between proteins or monomers, so we might expect homogeneous configurations; however, quasi-spherical protein clusters form, and essentially all proteins end up in a few large clusters in which several proteins interact with one DNA segment (Fig. 2; Movie S3). Again, almost no clustering occurs in the absence of any attraction between protein and DNA (Fig. 2*C*). Here, clusters are reminiscent of those seen when positively charged

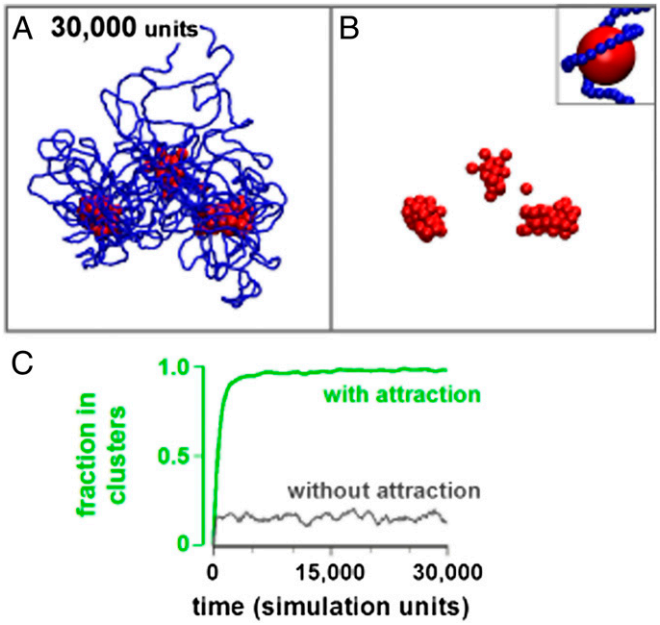

Fig. 2. Large proteins form quasi-spherical clusters on binding to DNA. MD simulations involving one string of blue beads (2.5 nm diameter) representing 73.53 kbp of DNA (persistence length, 50 nm; volume fraction, 0.16%; the radius of gyration of the unconfined polymer is ~645 nm) (18) interacting (12.4 k_BT,;range, 10 nm from center of protein bead to DNA bead) with 100 DNA-binding proteins (red spheres, 12.5 nm diameter; volume fraction 0.2% equivalent to 3.15 μM) in a cube (375 × 375 × 375 nm). (*A* and *B*) Two views (with/without DNA) of one structure after 30,000 simulation units; many complexes cluster. (*Inset*) Example of DNA wrapping reminiscent of that around a nucleosome core from a simulation involving exactly the same parameters but only one protein; such structures are rarely seen with many proteins. (*C*) The fraction of beads in clusters increases with time only if there is an attraction (two proteins are in one cluster if center-to-center distance is <17.5 nm).

Downloaded by guest on February 24, 2022

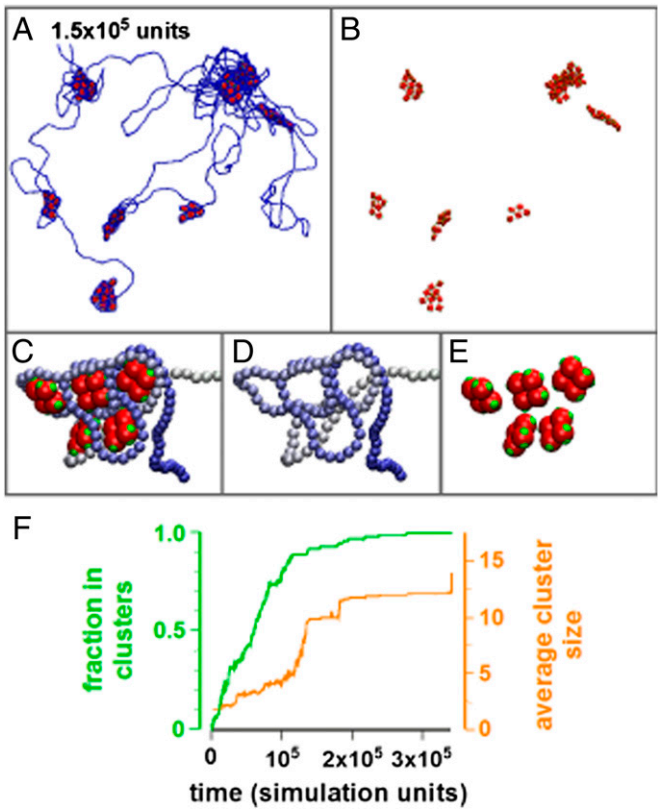

Fig. 3. Binding of nucleosome cores generates disordered chromatin fibers. MD simulations involving one string of blue beads (2.5 nm diameter) representing 36.76 kbp of DNA (persistence length, 50 nm; volume fraction, 0.02%; the radius of gyration of the unconfined polymer is ~456 nm) (18) interacting (energy, 4.31 k_BT; range, 3.5 nm from center of DNA bead to center of DNA-binding patch) with 100 nucleosomal cores (volume fraction 0.006%) in a cube (750 × 750 × 750 nm). Each core is represented by four planar (red) spheres each bearing two (green) binding sites. (*A* and *B*) Two views (with/without DNA) of one structure after 1.5×10^5 simulation units (one unit corresponds to 35 ns, assuming a viscosity of 1 cP); many cores cluster. (*C–E*) Three views (with/without DNA or cores) of one cluster in *A* shown from a different viewpoint; DNA is folded around cores much as in nucleosomes. (*F*) Both the fraction of beads in clusters and average cluster size increase with time (two bound cores form one cluster if center-to-center distance of each core is <17.5 nm).

nanoparticles (10–100 nm in diameter) are mixed in vitro with 166 kbp of DNA (14, 15). The histone-like wrapping found previously (20, 21) is only seen when DNA interacts with one protein (Fig. 2*B*, *Inset*).

To simulate nucleosome formation, we modeled DNA interacting with cylinders having eight DNA-binding patches on the circumference (Fig. S4*B*). On binding, the DNA can wrap around such cylinders (representing an octameric nucleosomal core), which again cluster as one segment contacts two nucleosomes (Fig. 3). Although strings of nucleosomes can form superbeads in vitro (22), there remains little evidence for such higher-order structures in vivo (23); moreover, our clusters stand out because they are separated in space by intervening DNA that is naked. Even so, the irregularly folded fiber a cluster is analogous to that found in current models for chromatin (23).

Large Proteins Form Quasi-Spherical Clusters on Chromatin Fibers. Cluster shape depends on polymer flexibility. We illustrate this with 20-nm proteins that bind nonspecifically to euchromatin, modeled as a string of 20-nm beads (2 kbp/bead). Then the ratio between monomer and protein diameter is the same as in Fig. 1. The fiber has a persistence length of 60 nm (24, 25) and is relatively more flexible than naked DNA (the ratio of persistence length to thickness is ~20 for naked DNA and ~3 for euchromatin). The protein could represent a complex containing RNA polymerase and some transcription factors; it is able to bind to two or more different DNA segments. Despite any explicit interaction between complexes, quasi-spherical clusters again form (Fig. 4; Movie S4), which are reminiscent of transcription factories (5, 6).

RNA Polymerase II and NFκB Cluster on Binding to Whole Human Chromosomes. Thus far, our proteins have equal affinities for all monomers in the fiber; we now consider proteins binding only to selected monomers. Here we model RNA polymerase II and one of its transcription factors (NFκB) binding to human chromosomes 5, 8, 14, and 17 modeled as 30-nm fibers (3 kbp/bead). These examples were chosen for several reasons. First, these two proteins are often the molecular ties that stabilize chromatin loops (5, 6, 26). Second, TNFα is a potent cytokine that signals through NFκB to orchestrate the inflammatory response. NFκB is normally cytoplasmic, but addition of TNFα to diploid (G0) human umbilical vein endothelial cells (HUVECs) induces phosphorylation of the p65 subunit of NFκB, nuclear import, and binding to thousands of sites around the genome; then, several hundred genes are up-/down-regulated as new intra/interchromosomal contacts appear. Here, we model the situation 30 min after adding TNFα, a time when we have detailed information on protein binding and how binding influences

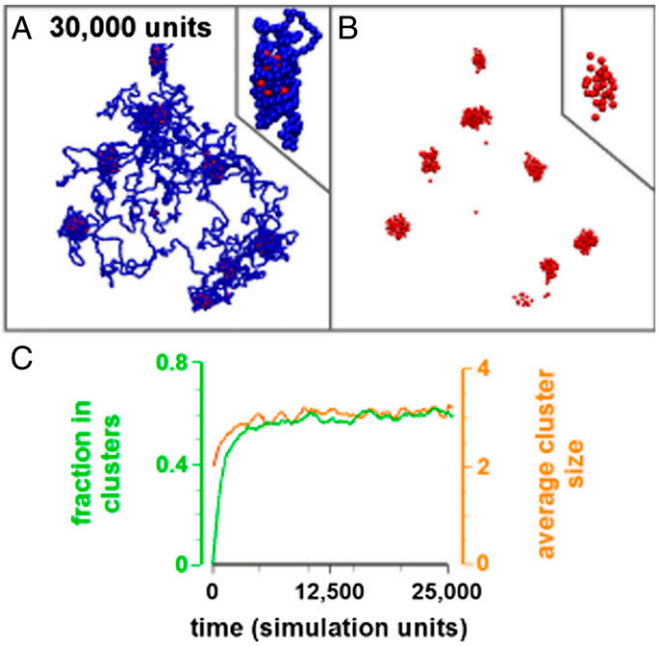

Fig. 4. Clustering of 20-nm complexes on binding to euchromatin. MD simulations involving one string of blue beads (20 nm diameter; 2 kbp of DNA) representing 10 Mbp euchromatin (persistence length, 60 nm; the radius of gyration of the unconfined polymer is ~1.4 μm) interacting (4.12 k_BT; range, 26 nm from DNA center to polymerase center) with 100 complexes containing RNA polymerases and transcription factors (red beads, 20 nm diameter) in a cube (2 × 2 × 2 μm). (*A* and *B*) Two views (with/without DNA) of one structure after 30,000 simulation units (one unit corresponds to 0.36 ms, assuming a nucleoplasmic viscosity of 10 cP); complexes cluster. (*Insets*) High-power views of one cluster. (*C*) Both the fraction of beads in clusters and average cluster size increase with time (two complexes form one cluster if center-to-center distance is <28 nm; using this stringent threshold, the fraction in clusters only reaches ~0.6, despite all but three red beads appearing to be in clusters in *B*).

Downloaded by guest on February 24, 2022

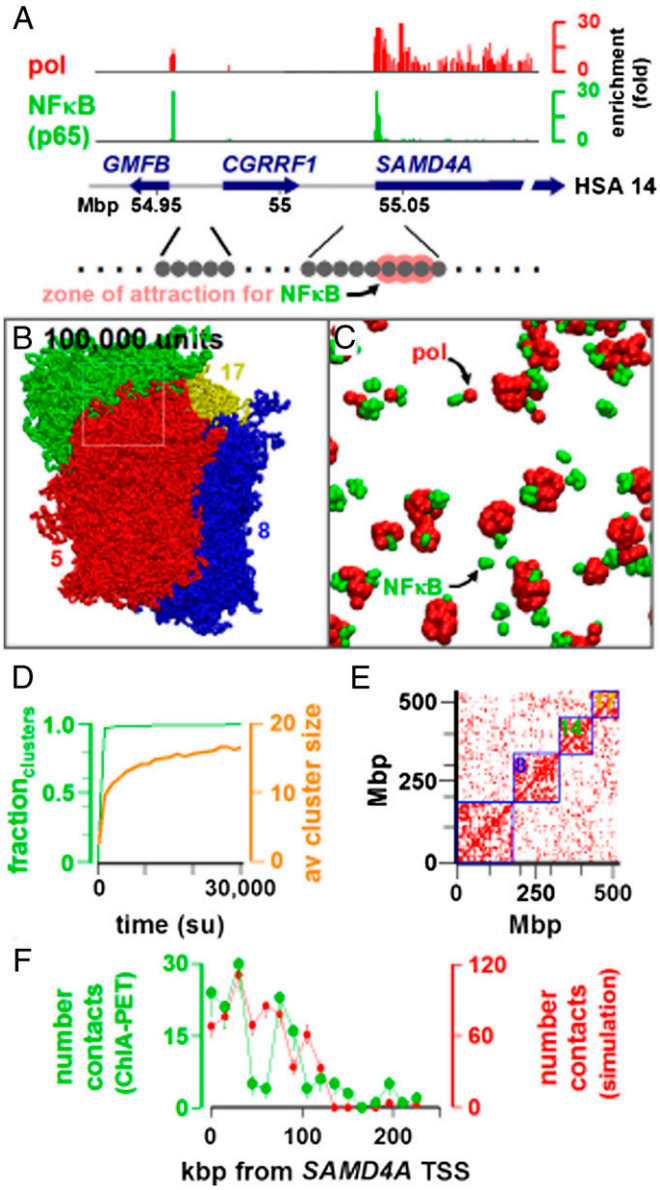

Fig. 5. Clustering of NFκB and RNA polymerase II bound to human chromosomes 5, 8, 14, and 17. MD simulations involving four strings of beads (diameter 30 nm) representing human chromosomes 5 (red), 8 (blue), 14 (green), and 17 (yellow) modeled as polymers of appropriate length (persistence length, 90 nm; volume fraction, 10%) in a cube (3 × 3 × 3 μm). Radii of gyration of unconfined chromosomes 5, 8, 14, and 17 are ∼7.4, ∼6.6, ∼5.6, and ∼4.9 μm, respectively, so polymers are in the semidilute to concentrated regime (18). The cube also contained 5,000 NFκB complexes (green 9-nm spheres) plus 5,000 RNA polymerases (red 18-nm spheres) that bind to cognate sites on the chromosomes. As a result, there are four types of beads: nonbinding, able to bind just NFκB or just the polymerase, and able to bind both. Binding data for p65 (a subunit of NFκB) and the polymerase were obtained by ChIP-seq using HUVECs 30 min after stimulation with TNFα. For polymer:polymerase interactions, the interaction energy was set to 15.95 k_BT and the range (between centers of DNA and protein beads) to 43.2 nm; for polymer:p65 interactions, corresponding values were 13.52 k_BT and 36 nm. (*A*) Browser views of the 5′ region of *SAMD4A* showing binding sites for the polymerase and p65 (fold enrichment indicated). The cartoon below the map indicates how binding of just NFκB to each 3-kbp segment is modeled; only the indicated 3 of 14 beads (at the *SAMD4A* promoter) possess a surrounding attractive zone (pink) and can bind NFκB. (*B*) A snapshot taken after 100,000 simulation units (equivalent to ∼120 s, assuming a viscosity of 10 cP). (*C*) Magnification of *Inset* in *B* without chromosomes to highlight protein clustering. (*D*) Both the fraction of beads in clusters and average cluster size increase with time (two polymerases or two NFκB complexes form one cluster if center-to-center distance is <36 nm). (*E*) Contacts (marked as a cross and defined as center-to-center distance <90 nm) within and between the four chromosomes; the four remain segregated in territories to form more intra- than interchromosomal contacts (indicated by the high concentration of crosses in blue boxes). (*F*) Simulations and ChIA-PET yield similar contacts. Data on contacts made by every 3-kbp region within *SAMD4A* were obtained from the simulation (contact defined as two monomers lying within 90 nm) or ChIA-PET (using data from ref. 16; contact defined as number of paired reads with no base pair mismatch in the *SAMD4A* tag and up to two mismatches in the paired tag). The contact number (coarse-grained into 15-kbp bins) detected by the two methods falls in much the same way with distance from the transcription start site (TSS).

transcriptional activity and inter/intrachromosomal contacts (16, 27, 28).

To make our model more realistic, the number of monomers in each polymer reflects chromosome length. Using ChIP-seq data (obtained using antibodies targeting the polymerase or p65) (16), each 3-kbp monomer is categorized as able to bind (or not) the polymerase and/or NFκB. For example, the three beads at the promoter of one gene responding to TNFα, sterile alpha motif domain containing 4A (*SAMD4A*), bind both proteins, whereas their neighbors do not (Fig. 5*A*). All four polymers are confined in a cube containing 5,000 polymerases and 5,000 NFκB molecules (equivalent to micromolar concentrations). Once again, no protein has any explicit affinity for another protein. These large-scale (parallel) simulations involve 181,954 monomers and at least 10^8 time steps (taking 1 wk on 24 processors).

As might be expected, sequence-specific binding also drives aggregation, with the polymerase and transcription factor often becoming concentrated at different ends of a cluster (Fig. 5 *B–D*); we attribute this to NFκB binding to beads at a promoter next to others binding the polymerase on the gene body. As before, clustering is accompanied by a decrease in pairing energy, and it does not occur in the absence of a protein:DNA interaction. Similar clusters form in simulations involving only NFκB that binds with one of six different affinities to cognate beads (i.e., where the attraction reflects the peak height seen by ChIP-seq; Fig. S7).

At the global level, a segment within each of the four chromosomes tends to contact another part of the same chromosome rather than another chromosome; in the resulting contact map, the four blue squares contain higher densities of contacts compared with other areas (Fig. 5*E*). Similar maps are obtained using Hi-C, where intrachromosomal contacts within individual chromosome territories also predominate (29). However, our contacts depend strongly on starting conditions. Thus, the simulation in Fig. 5 began with a self-avoiding conformation in each quarter of the confining cube, and a similar outcome is obtained when starting with four similarly placed mitotic-like structures (30) (*Methods*). In contrast, starting with four intermingled self-avoiding random walks (*Methods*) yields little de-mixing and no evidence for territory formation. This result is consistent with the reptation dynamics of long polymers being slow relative to simulation times (24, 30). Nevertheless, these different initial conditions all give protein clusters of roughly equal size.

At the local level, contacts made by different segments within *SAMD4A* reflect those seen in vivo. Before stimulation with TNFα, *SAMD4A* is not transcribed, no NFκB or polymerase binds, and it contacts few other chromosomes. However, 30 min after adding the cytokine (the situation modeled), NFκB binds to the promoter, pioneering polymerases now transcribe the first half of this 221-kbp gene (Fig. 5*A*) (27), and ChIA-PET reveals that this transcribed half contacts many other segments on the same and other chromosomes; these also tend to bind the polymerase and/or NFκB (16, 28). The contacts seen in the simulation mirror those detected by ChIA-PET (Fig. 5*F*). At the global level, a typical 3-kbp segment/bead within 18 other up-regulated genes

Downloaded by guest on February 24, 2022

(which also bind both the polymerase and NFκB) contacts more other segments/beads than an average (nonbinding) bead, and these contacts tend to be with other protein-binding regions (Table S1, compare row 3 with rows 1 and 4). These results confirm that DNA segments binding the polymerase/NFκB cluster in the simulations and that the resulting contacts reflect those seen by ChIA-PET.

Discussion

Our MD simulations uncover an emergent property of a system involving proteins that bind to, and dissociate from, DNA; on binding, the proteins spontaneously cluster, even in the absence of any explicit interaction between proteins or monomers, and this clustering inevitably organizes the genome. This phenomenon is generic, robust, and occurs over a range of conditions (e.g., proteins/monomers of 2.5–30 nm, 1–100 μM protein concentration, 0.26–9.7% volume fraction of DNA, interaction energies in the range sufficient for binding and dissociation, and when proteins are initially prebound randomly along the DNA). Although our simulations might not reach a true equilibrium or global energy minimum, clusters are nevertheless stable over long periods.

What forces might drive aggregation in the absence of any explicitly included interaction between proteins or monomers? Several physical mechanisms are illustrated in Fig. 6. In all cases where significant clustering is seen, proteins bind to at least two different DNA monomers to form a molecular bridge. Such bridging distorts the DNA locally, for example, by bringing two distant segments together (Figs. 1–4), straightening or bending the DNA (Figs. 1 and 2), or creating loops at many scales (Fig. 5*E*). These distortions have associated entropic and bending penalties. Then, rather than creating such distortions in several places along DNA, it seems energetically advantageous to group together the unfavorable conformations (Fig. 6 *A*–*C*). For example, two bridges between the same two DNA segments do not exact as large an entropic cost as bridging two different segments (Fig. 6 *B* and *C*). Kinetic effects may also contribute to clustering. For instance, bridge formation enhances the local DNA concentration, which can then more effectively sieve out unbound proteins (Fig. 6*D*; Movie S2). Once a cluster has formed and a protein dissociates, it is likely to rebind to the same cluster simply because the local concentration of binding sites is so high. Here, the protein concentration in the cluster is maintained despite the homogenizing effects of dissociation and diffusion (as is seen with the LacI protein in bacteria) (31). In addition, if segments between bridges are coaligned, then, when a protein dissociates, it is likely to rebind close by, so the bridges zip together (Movie S2). We dub the combination of these effects (Fig. 6) the "bridging-induced attraction."

Protein binding is required (Fig. S3), but is bridging required? One might imagine that binding without bridging would also locally distort DNA, and it would be energetically advantageous to cluster the resulting distortions together (Fig. S8*A*). Moreover, theory shows that protein binding can locally alter the persistence length and promote elastic interactions between proteins to affect force-extension measurements in single-molecule experiments (32, 33). However, monovalent proteins show very little clustering in our simulations (Figs. S5 and S6), presumably because any local distortion caused is too small to lead to clustering. We note that previous theory (33) predicts attraction only when the DNA is under tension (although the effect may become more important if binding twists DNA, which is not considered here). We also note that bound (but not bridging) proteins can snag as DNA segments slide past each other (Fig. S8*B*); this effect is again insufficient to lead to clustering, presumably because snagged entanglements are quickly eliminated at the polymer concentrations considered here. Therefore, we conclude that bridging is required to obtain clustering.

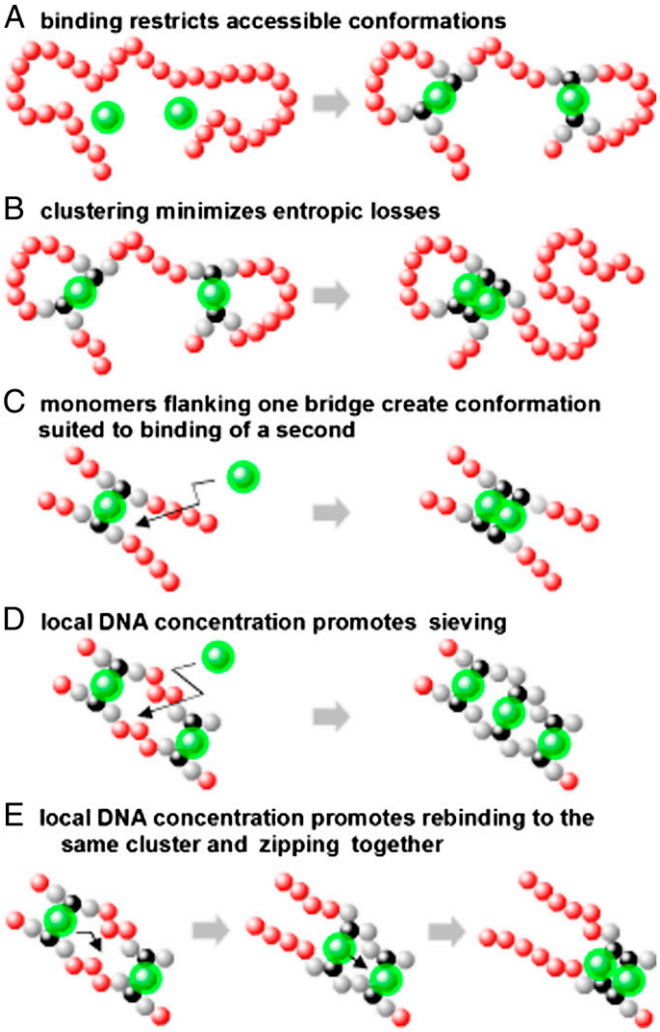

Fig. 6. The bridging-induced attraction. Each panel illustrates part of a long polymer and some DNA-binding proteins (green spheres surrounded by attractive zones). (*A*) When proteins bind, the mobility of black monomers is restricted (reducing their entropy); the gray flanking monomers also lose some entropy, and more distant ones progressively less (not indicated). (*B*) The two structures contain the same number of monomers and proteins, but the one on the right will be more stable as it contains one less loop and fewer gray monomers. (*C*) Once one bridge forms, monomers on each side of the bridge are likely to be positioned in a way favoring binding of a second bridge. This binding does not exact the full entropic cost, as much of that cost was paid when the first bridge formed. With stiff polymers like naked DNA, this probably drives the formation of rows of bound proteins (Fig. 1; Fig. S5). (*D*) Once two bridges connect two DNA segments, the resulting high local concentration creates a collisional cross section likely to sieve out any protein that diffuses by. (*E*) When the left-hand bridge dissociates, rebinding nearby is promoted by the high local DNA concentration. After several steps of dissociation/rebinding, or sliding, the resulting zipping together gives the structure on the right, which is the most stable (with four fewer gray monomers than the other two structures).

What determines cluster shape? It depends particularly on protein size and polymer persistence length. Thus, for 2.5-nm proteins binding to naked DNA, the polymer is stiff on the length scale of the protein, so little bending is induced and rows form (Fig. 1), like those seen on H-NS binding (1, 2, 13). For 10-nm proteins, DNA is more likely to bend around the proteins, and quasi-spherical aggregates result (Fig. 2), as seen when DNA binds to nanoparticles (14, 15). For still-larger complexes (representing RNA polymerase II and transcription factors) binding

PNAS PLUS
BIOPHYSICS AND COMPUTATIONAL BIOLOGY
PHYSICS
Downloaded by guest on February 24, 2022

to chromatin, the polymer is flexible enough to wrap around the complexes, and quasi-spherical clusters again form (Figs. 4 and 5; Fig. S7), which are analogous to transcription factories (5, 6). Although these clusters contain DNA in the interior, unlike those found in vivo (6), they do yield the same patterns of local (genic) and multiscale (intrachromosomal) contacts seen in living cells (Fig. 5; Fig. S7).

Note that the (entropic) depletion attraction (19) cannot drive clustering. Although it induces macromolecular clustering in crowded environments (7, 34), it requires protein (or DNA) concentrations 10- to 100-fold higher (i.e., 20–30% by volume) than used in Figs. 1–4; if it acted, our proteins would cluster in the absence of any protein:DNA interaction, but they do not (Fig. S3). Furthermore, the bridging-induced attraction works in regimes in which the DNA is dilute (so its radius of gyration is the same order of magnitude as the size of the simulation cube; Figs. 3 and 4), semidilute (Figs. 1 and 2), and at higher concentrations where the chromatin concentration is roughly that found in vivo (Fig. 5; Fig. S2). Of course, in a living cell, the classical depletion attraction acting through many crowding macromolecules (absent in our simulations) will augment the bridging-induced attraction to further promote clustering.

The compaction and clustering observed here share similarities with other phenomena encountered in polymer science. For example, when polymers are mixed with large and charged colloids, the former can wrap around the latter (Fig. 2*B*, *Inset*) (20, 21). When polymers are mixed with smaller charged colloids, the colloids can form bridges that stabilize loops (35). A recent theoretical treatment considers such bridge-forming colloids and finds the most energetically favorable conformation to be the one where colloid bridges fold the polymer into small loops that can then slide along other segments of the polymer (35). Our simulations extend these findings to proteins and sequence-specific binding and allow examination of important metastable states that are not accessible to equilibrium theories. Another system involves polymer-stabilized colloidal dispersions (36) or dispersions in liquid crystals (37); in the latter case, minimizing elastic distortions in the host medium induces a flocculation analogous to our cluster formation. Polycations can also induce DNA condensation through a process known as disproportionation (38); partially neutralized DNA segments aggregate into droplets (driven by short-range and Coulombic forces), surrounded by a halo of negatively charged segments (39–41). Although we do not consider Coulombic interactions, the principles are related: interstrand attraction is mediated by bridging polycations (which compacts DNA) as configurational entropy of unbound regions is maximized. This work is closely related to the general phenomenon of polyelectrolyte-mediated interactions between like-charged objects (42). Another phenomenon involves protein-induced DNA bending that reduces the radius of gyration, as well as unexpectedly increasing protein-DNA affinity (43).

In conclusion, we suggest that protein binding induces a bridging attraction that drives protein clustering and genome reorganization. Then, the system must either spend energy to prevent it, or, as seems likely, it goes with the flow and uses it. In the specific case of complexes containing polymerases and replication/transcription factors, we suggest this attraction drives the formation of the factories that carry out the vital processes of replication and transcription (5, 6, 44). We also expect that improvements in computation and high-throughput sequencing will soon allow detailed comparison of the intrachromosomal contact maps seen in simulations and cells (e.g., after choosing interaction energies between additional proteins and chromosomes; Fig. S7).

Methods

Brownian dynamics (BD) simulations were run with large-scale atomic/molecular massively parallel simulator (LAMMPS) code, used in the BD mode (i.e., with an MD algorithm with a stochastic thermostat) (45). DNA and chromatin are modeled as bead-and-spring polymers using finite extensible nonlinear elastic (FENE) bonds (maximum extension 1.6 times bead diameter) and a bending potential that allows persistence length to be set. Protein:protein and template:template interactions involve only steric repulsion. For template:protein interactions, all parts of a protein, or only sticky patches within one, are uniformly attracted to the template. All participants are confined within a cube with periodic boundary conditions, but strings are unwrapped for presentational purposes (i.e., disconnected strings are rejoined). Parameters are listed in the figure legends and/or in Table S2 and *SI Methods* (which provides details on the simulations, including the force field used).

ACKNOWLEDGMENTS. C.A.B. is supported by Engineering and Physical Sciences Research Council Grant EP/I034661/1, and A.P. is supported by the Biotechnology and Biological Sciences Research Council and the ERASysBio+ initiative under the FP7/ERA-NET Plus scheme. P.R.C. is an E. P. Abraham Professor of Cell Biology and a Professorial Fellow of Lincoln College.

1. Wiggins PA, Dame RT, Noom MC, Wuite GJ (2009) Protein-mediated molecular bridging: A key mechanism in biopolymer organization. *Biophys J* 97(7):1997–2003.
2. Wang W, Li GW, Chen C, Xie XS, Zhuang X (2011) Chromosome organization by a nucleoid-associated protein in live bacteria. *Science* 333(6048):1445–1449.
3. Tark-Dame M, Luijsterburg MS, Heermann DW, van Driel R (2012) Understanding genome function: Quantitative modelling of chromatin folding and chromatin-associated processes. *Genome Organization and Function in the Nucleus*, ed Rippe K (Wiley, Winheim, Germany). pp. 535–555.
4. Biggin MD (2011) Animal transcription networks as highly connected, quantitative continua. *Dev Cell* 21(4):611–626.
5. Chakalova L, Fraser P (2010) Organization of transcription. *Cold Spring Harb Perspect Biol* 2(9):a000729.
6. Papantonis A, Cook PR (2013) Transcription factories: Genome organization and gene regulation [published online ahead of print April 18, 2013]. *Chem Rev*, doi:10.1021/cr300513p.
7. Marenduzzo D, Micheletti C, Cook PR (2006) Entropy-driven genome organization. *Biophys J* 90(10):3712–3721.
8. Nicodemi M, Panning B, Prisco A (2008) The colocalization transition of homologous chromosomes at meiosis. *Phys Rev E Stat Nonlin Soft Matter Phys* 77 (6 Pt 1):061913.
9. de Vries R (2011) Influence of mobile DNA-protein-DNA bridges on DNA configurations: Coarse-grained Monte-Carlo simulations. *J Chem Phys* 135(12):125104.
10. Kalhor R, Tjong H, Jayathilaka N, Alber F, Chen L (2012) Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat Biotechnol* 30(1):90–98.
11. Barbieri M, et al. (2012) Complexity of chromatin folding is captured by the strings and binders switch model. *Proc Natl Acad Sci USA* 109(40):16173–16178.
12. Fritsche M, Li S, Heermann DW, Wiggins PA (2012) A model for Escherichia coli chromosome packaging supports transcription factor-induced DNA domain formation. *Nucleic Acids Res* 40(3):972–980.
13. Dame RT, Noom MC, Wuite GJ (2006) Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation. *Nature* 444(7117):387–390.
14. Zinchenko AA, Yoshikawa K, Baigl D (2005) Compaction of single-chain DNA by histone-inspired nanoparticles. *Phys Rev Lett* 95(22):228101.
15. Zinchenko AA, Sakaue T, Araki S, Yoshikawa K, Baigl D (2007) Single-chain compaction of long duplex DNA by cationic nanoparticles: Modes of interaction and comparison with chromatin. *J Phys Chem B* 111(11):3019–3031.
16. Papantonis A, et al. (2012) TNFα signals through specialized factories where responsive coding and miRNA genes are transcribed. *EMBO J* 31(23):4404–4414.
17. Fullwood MJ, et al. (2009) An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* 462(7269):58–64.
18. Doi M, Edwards SF (1986) *The Theory of Polymer Dynamics* (Clarendon Press, Oxford, UK).
19. Asakura S, Oosawa F (1958) Interactions between particles suspended in solutions of macromolecules. *J Polym Sci* 33(126):183–192.
20. Sakaue T, Yoshikawa K, Yoshimura SH, Takeyasu K (2001) Histone core slips along DNA and prefers positioning at the chain end. *Phys Rev Lett* 87(7):078105.
21. Gurovitch E, Sens P (1999) Adsorption of polyelectrolyte onto a colloid of opposite charge. *Phys Rev Lett* 82(2):339–342.
22. Zhao H, et al. (1999) The structure of the nucleosome core particle of chromatin in chicken erythrocytes visualized by using atomic force microscopy. *Cell Res* 9(4):255–260.
23. Hansen JC (2012) Human mitotic chromosome structure: What happened to the 30-nm fibre? *EMBO J* 31(7):1621–1623.
24. Cook PR, Marenduzzo D (2009) Entropic organization of interphase chromosomes. *J Cell Biol* 186(6):825–834.
25. Langowski J (2006) Polymer chain models of DNA and chromatin. *Eur Phys J E Soft Matter* 19(3):241–249.
26. Li G, et al. (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* 148(1–2):84–98.
27. Wada Y, et al. (2009) Visualizing a wave of transcription as it sweeps along activated human genes. *Proc Natl Acad Sci USA* 106:18357–18361.

Downloaded by guest on February 24, 2022

28. Papantonis A, et al. (2010) Active RNA polymerases: Mobile or immobile molecular machines? *PLoS Biol* 8(7):e1000419.
29. Lieberman-Aiden E, et al. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326(5950):289–293.
30. Rosa A, Everaers R (2008) Structure and dynamics of interphase chromosomes. *PLOS Comput Biol* 4(8):e1000153.
31. Kuhlman TE, Cox EC (2012) Gene location and DNA density determine transcription factor distributions in Escherichia coli. *Mol Syst Biol* 8:610.
32. Yan J, Marko JF (2003) Effects of DNA-distorting proteins on DNA elastic response. *Phys Rev E Stat Nonlin Soft Matter Phys* 68(1 Pt 1):011905.
33. Rudnick J, Bruinsma R (1999) DNA-protein cooperative binding through variable-range elastic coupling. *Biophys J* 76(4):1725–1733.
34. Marenduzzo D, Finan K, Cook PR (2006) The depletion attraction: An underappreciated force driving cellular organization. *J Cell Biol* 175(5):681–686.
35. Baulin VA, Johner A, Avalos JB (2010) Aggregation of amphiphilic polymers in the presence of adhesive small colloidal particles. *J Chem Phys* 133(17):174905.
36. Gast AP, Leibler L (1986) Interactions of sterically stabilized particles suspended in a polymer solution. *Macromolecules* 19(13):686–691.
37. Galatola P, Fournier JB, Stark H (2003) Interaction and flocculation of spherical colloids wetted by a surface-induced corona of paranematic order. *Phys Rev E Stat Nonlin Soft Matter Phys* 67(3 Pt 1):031404.
38. Zhang R, Shklovskii BI (2005) Phase diagram of solution of oppositely charged polyelectrolytes. *Physica A* 352(1):216–238.
39. Neu M, Fischer D, Kissel T (2005) Recent advances in rational gene transfer vector design based on poly(ethylene imine) and its derivatives. *J Gene Med* 7(8):992–1009.
40. Ziebarth J, Wang Y (2009) Molecular dynamics simulations of DNA-polycation complex formation. *Biophys J* 97(7):1971–1983.
41. Kabanov VA, et al. (2000) Interpolyelectrolyte complexes formed by DNA and astramol poly(propylene imine) dendrimers. *Macromolecules* 33(26):9587–9593.
42. Podgornik R (2004) Polyelectrolyte-mediated bridging interactions. *J Polym Sci B* 42(19):3539–3556.
43. Medalion S, Rabin Y (2012) On binding of DNA-bending proteins to DNA minicircles. *J Chem Phys* 136(2):025102.
44. Cook PR (1999) The organization of replication and transcription. *Science* 284(5421): 1790–1795.
45. Plimpton S (1995) Fast parallel algorithms for short-range molecular dynamics. *J Comput Phys* 117(1):1–19.

Downloaded by guest on February 24, 2022