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.

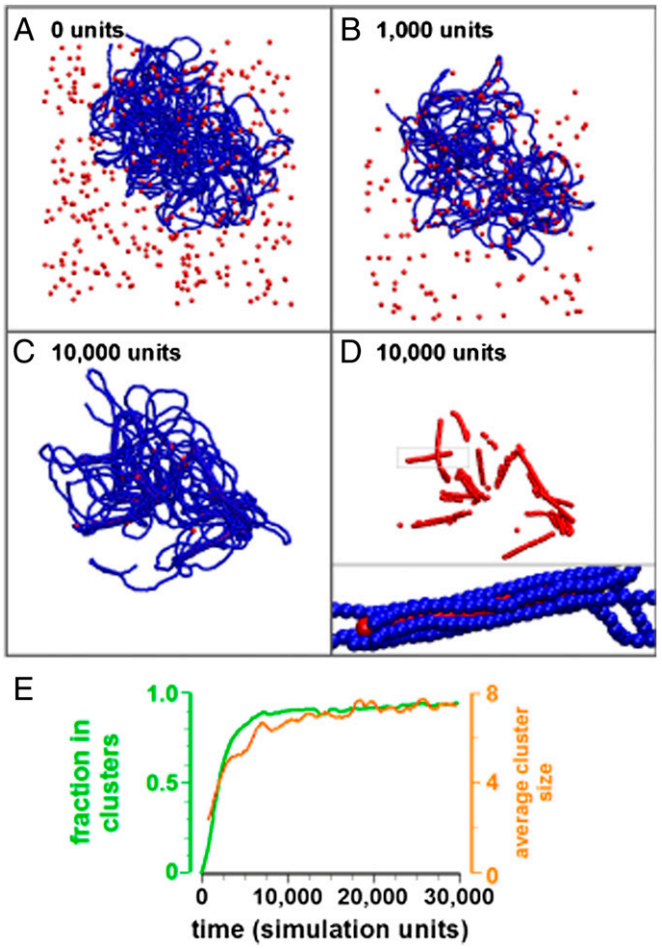


Fig. 1. Small proteins bind to DNA and form rows. MD simulations involving one string of blue beads (2.5 nm diameter) representing 36.75 kbp of DNA (persistence length, 50 nm; volume fraction, 0.26%; the radius of gyration of the unconfined polymer is ~456 nm) (18) interacting with 400 DNA-binding proteins (2.5-nm-diameter red spheres; volume fraction, 0.02%) in a cube ($250 \times 250 \times 250$ 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

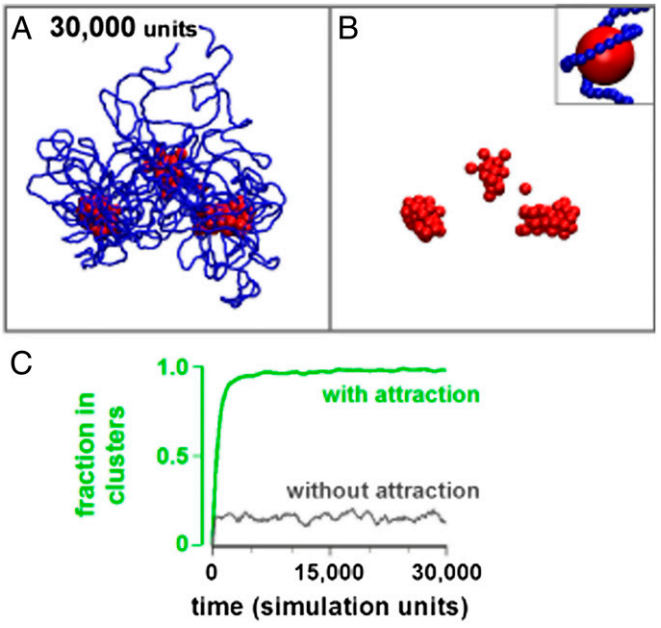


Fig. 2. Large proteins form quasi-spherical clusters on binding to DNA. MD simulations involving one string of blue beads (2.5 nm diameter) representing 73.53 kbp of DNA (persistence length, 50 nm; volume fraction, 0.16%; the radius of gyration of the unconfined polymer is ~645 nm) (18) interacting (12.4 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 ($375 \times 375 \times 375$ 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).

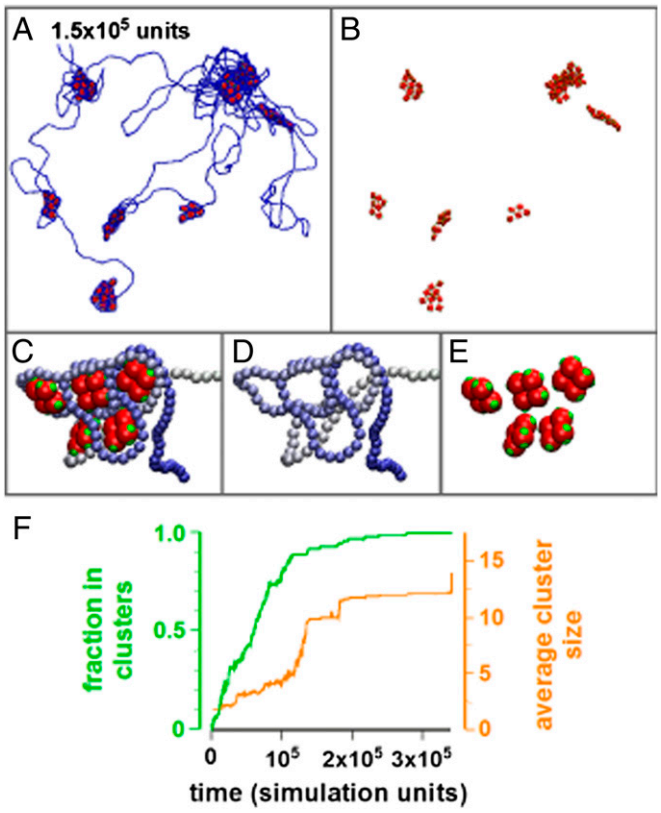


Fig. 3. Binding of nucleosome cores generates disordered chromatin fibers. MD simulations involving one string of blue beads (2.5 nm diameter) representing 36.76 kbp of DNA (persistence length, 50 nm; volume fraction, 0.02%; the radius of gyration of the unconfined polymer is ~456 nm) (18) interacting (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

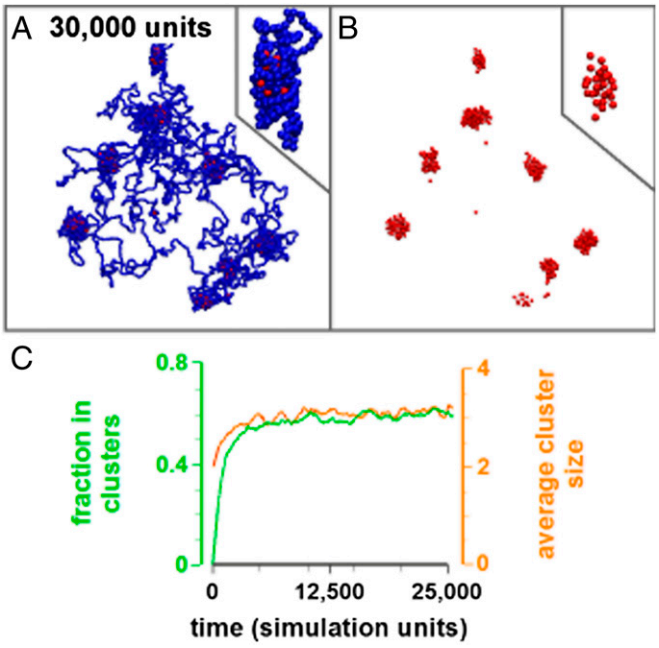


Fig. 4. Clustering of 20-nm complexes on binding to euchromatin. MD simulations involving one string of blue beads (20 nm diameter; 2 kbp of DNA) representing 10 Mbp euchromatin (persistence length, 60 nm; the radius of gyration of the unconfined polymer is ~1.4 μm) interacting (4.12 k_BT; range, 26 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*).

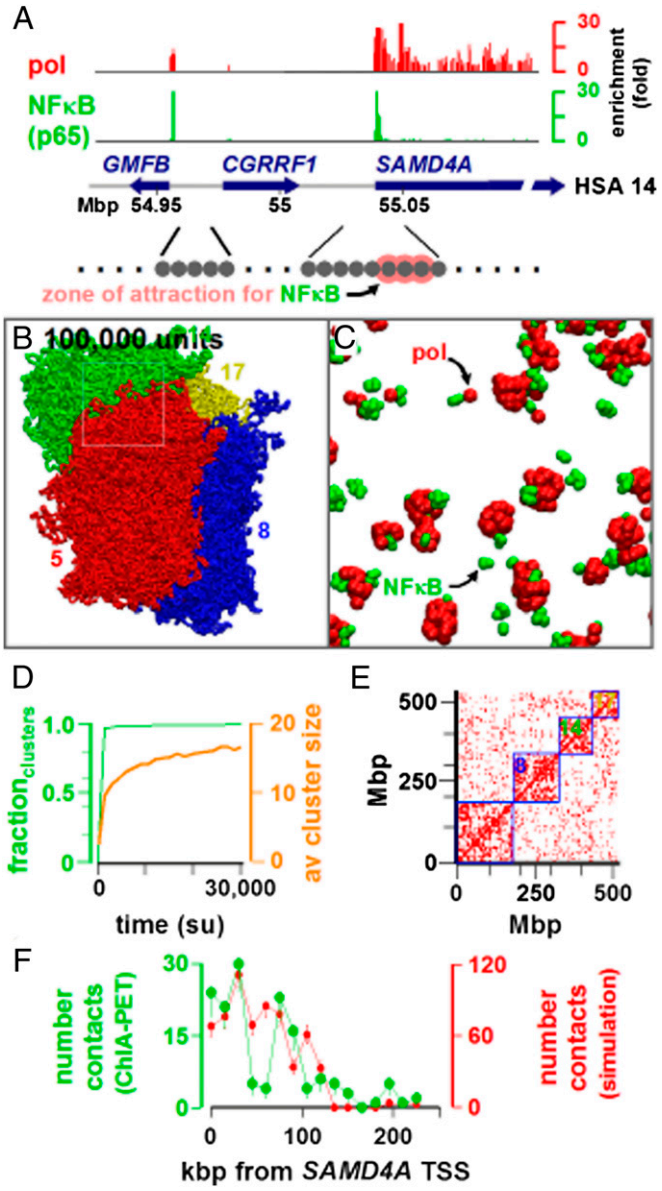


Fig. 5. Clustering of NFκB and RNA polymerase II bound to human chromosomes 5, 8, 14, and 17. MD simulations involving four strings of beads (diameter 30 nm) representing human chromosomes 5 (red), 8 (blue), 14 (green), and 17 (yellow) modeled as polymers of appropriate length (persistence length, 90 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

(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.

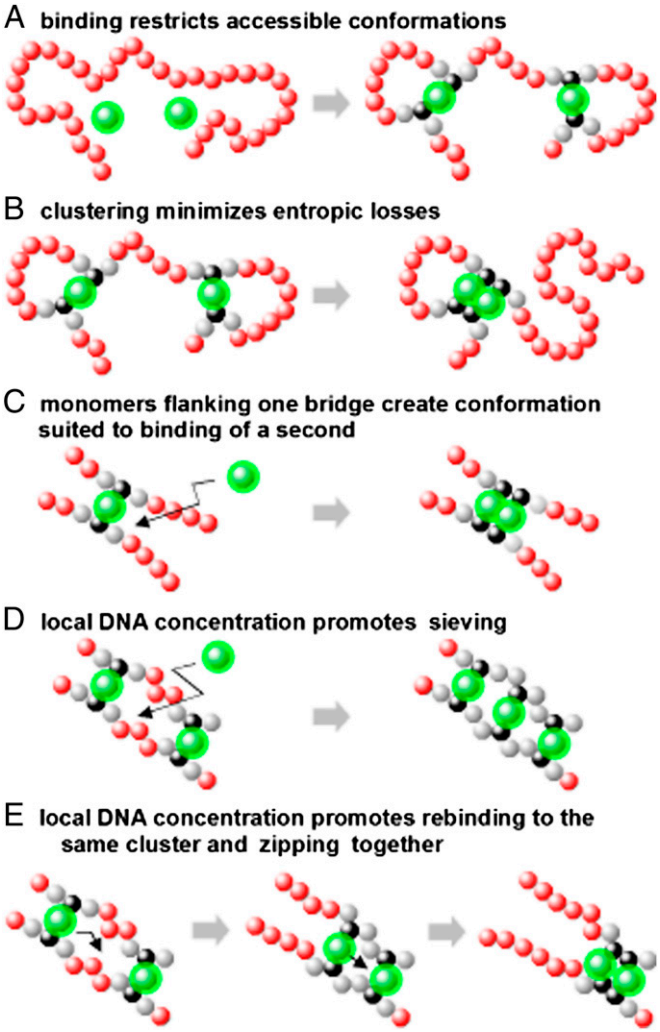


Fig. 6. The bridging-induced attraction. Each panel illustrates part of a long polymer and some DNA-binding proteins (green spheres surrounded by attractive zones). (*A*) When proteins bind, the mobility of black monomers is restricted (reducing their entropy); the gray flanking monomers also lose some 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

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.

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.

Supporting Information

Brackley et al. 10.1073/pnas.1302950110

SI Methods

Brownian Dynamics Simulations for DNA and DNA-Binding Proteins: Force Field. Brownian dynamics (BD) simulations were run with large-scale atomic/molecular massively parallel simulator (LAMMPS) code, used in the BD mode [i.e., with a molecular dynamics (MD) algorithm with a stochastic thermostat (1)]. Each bead, say the ith one, which could be either part of a protein or DNA, obeys a Langevin equation

$$m\frac{d^2\mathbf{x}_i}{dt^2} = -\nabla_i V - \gamma\frac{d\mathbf{x}_i}{dt} + \sqrt{2k_BT\gamma}\boldsymbol{\eta}(t), \qquad \text{[S1]}$$

where m is the mass ($m = 1$ unless otherwise specified), $\mathbf{x}_i$ is bead position, γ is the friction (typically $\gamma = 0.5$), k_B is the Boltzmann constant, T is the temperature, and $\boldsymbol{\eta}$ is a vector that represents random uncorrelated noise such that

$$\langle\boldsymbol{\eta}(t)\rangle = 0; \quad \langle\eta_\alpha(t)\eta_\beta(t')\rangle = \delta_{\alpha\beta}\delta(t-t'), \qquad \text{[S2]}$$

where the first delta is a Kronecker one (α and β denote Cartesian coordinates x, y, and z) and the second is a Dirac delta. The diffusion coefficient of the bead, D, is linked to previously defined quantities through $D = k_BT/\gamma$.

The interaction potential between DNA beads consists of three contributions, all standard in biopolymer physics. First, there is a truncated and shifted Lennard-Jones potential, acting between any two DNA beads, say the ith and jth, as follows:

$$V_{LJ}(r_{ij}) = \begin{cases} 4\varepsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right] + \varepsilon & \text{if } r_{ij} < 2^{1/6}\sigma \\ 0 & \text{otherwise,} \end{cases} \qquad \text{[S3]}$$

where $\varepsilon = k_BT$ is an energy scale, σ is the diameter of the DNA bead, and r_{ij} is the distance between the ith and jth beads. Eq. **S3** leads to a purely repulsive interaction and is also known as the Weeks-Chandler-Anderson potential. Second, there is a potential (V_{FENE}) between neighboring monomers so they are bonded by finitely extensible nonlinear elastic (FENE) springs

$$V_{FENE}(r = |\mathbf{r_{i+1}} - \mathbf{r_i}|) = -\frac{K_{FENE}R_0^2}{2}\log\left[1 - \left(\frac{r}{R_0}\right)^2\right], \qquad \text{[S4]}$$

where $\mathbf{x}_i$ and $\mathbf{x}_{i+1}$ denote the positions of monomers that are nearest neighbors along the chain, r is their distance, R_0 is the maximum extent of the bond (which we take equal to 1.6σ), and $K_{FENE} = 30k_BT/\sigma^2$ is the strength of the FENE springs. [Note that the combination between the FENE potential and the repulsive Weeks-Chandler-Anderson potential (Eq. **S2**) leads to a bond length approximately equal to σ.] Third, there is a bending potential defined in terms of angle between any three neighboring beads, θ, as follows:

$$V_{bending} = K_b[1 + \cos(\theta)], \qquad \text{[S5]}$$

where $K_b/(k_BT)$ determines the persistence length (in units of σ).

Protein:protein interactions are taken as purely repulsive (i.e., due solely to steric repulsion) using the same potential in Eq. **S3**, with the same parameters except for σ, which is replaced by the diameter of the protein (which varies as specified) or by the average between the diameters of interacting proteins (for patchy proteins or polymerase/NFκB mixtures).

For DNA:protein interactions, either all parts of a protein, or only sticky patches within one, are uniformly attracted to DNA. In the former case, the protein interacts with DNA beads through the following potential (a and b refer, for instance, to protein and DNA bead, respectively, whereas r denotes distance between bead centers)

$$V_{LJ}^{ab}(r_{ij}) = \begin{cases} 4\varepsilon_{ab}\left[\left(\frac{\sigma_{ab}}{r}\right)^{12} - \left(\frac{\sigma_{ab}}{r}\right)^{6} - \left(\frac{\sigma_{ab}}{r_{thr}}\right)^{12} + \left(\frac{\sigma_{ab}}{r_{thr}}\right)^{6}\right] & \text{if } r < r_{thr} \\ 0 & \text{otherwise,} \end{cases} \qquad \text{[S6]}$$

where ε_{ab} controls the magnitude of the DNA:protein interaction (together with the cutoff distance r_{thr}), and σ_{ab} is the average of the protein diameter and DNA thickness. The quantity r_{thr} is the interaction range, above which particles stop interacting. See Table S2 for parameters included in Eq. **S6**.

All participants are confined within a cube with periodic boundary conditions (so when a monomer or protein exits through a surrounding wall, it reenters through the opposite one), but strings are unwrapped for presentational purposes (i.e., disconnected strings are rejoined). Snapshots from BD simulations were prepared via the Visual Molecular Dynamics software (2).

Modeling DNA Bridges and Histone-Like Proteins as Patchy Particles. The potential function in Eq. **S6** permits multiple contacts between a protein and DNA, in contrast to many transcription factors (and other DNA-associated proteins such as H-NS) that bind to DNA through only two sites. We therefore used a refined model where the protein is a patchy particle that can only interact with DNA through localized areas on the surface. In practice, a patchy protein is modeled as a collection of spheres; one represents the whole protein (diameter 1σ), whereas the two others (with diameters of 0.2σ), lie within the largest and touch its north and south poles (so no parts of the small spheres protrude from the surface of the largest; Fig. S4*A*). However, an attractive field—defined as in Eq. **S6**—extends 0.1σ away from the surface of both small spheres. The collection of such spheres is then treated as one rigid body in the BD scheme that rotates and translates as a single entity. All beads making up the patchy particle interact with other beads in the simulations through Eq. **S6**, where the strength and cutoff of the interactions depend on bead type. As a result, the large sphere in the patchy particle now interacts only through steric repulsion with other proteins and DNA, whereas the small spheres making up the patches are attracted solely to DNA (their only interaction with other proteins involves steric repulsion).

The nucleosome core contains pairs of four histones; 146 bp of DNA is usually wrapped around the circumference of this octamer. We model the core as a disk (~10.5 nm diameter, ~6 nm width) in which four large spheres (diameter 2σ, ~5 nm) are arranged in a square with centers 2σ apart (Fig. S4*B*). These spheres have only a steric interaction with DNA. Eight smaller spheres (diameter 1σ) are partially contained within the larger spheres, and attractive fields extend 0.25σ from their surfaces. As a result, the smaller spheres again act as binding sites that interact through Eq. **S6** with DNA via a Lennard-Jones potential truncated at 1.4σ.

Modeling Four Human Chromosomes. We model binding of RNA polymerase II and NFκB to human chromosomes 5, 8, 14, and 17. Each polymer contains the appropriate number of 30-nm beads (each representing 3 kbp) that reflect chromosome length. Each bead is categorized as able to bind the polymerase and/or NFκB—in particular, the p65 subunit—with the interaction energies indicated in Table S2. Published ChIP-seq data (coordinates based on hg18), obtained using human umbilical vein endothelial cells (HUVECs) 30 min after stimulation with TNFα (3), were analyzed. Centers of peaks were first identified using PeakRanger (4) using default parameters. Then, each 200-bp window of the genome was classified as binding or not based on whether it overlapped an enriched region, and coordinates were converted to hg19. For Fig. 5, a 3,000-bp bead was classified as binding if any 200-bp window within it was classified as binding, and all four polymers are confined in a cube ($3 \times 3 \times 3$ μm) containing 5,000 polymerases and 5,000 NFκB molecules (with diameters of 18 and 9 nm, respectively). For Fig. S7, monomers were given one of six different affinities for NFκB (with interaction energies ranging from 0 to 13.52 k_BT) depending on peak height in the ChIP coupled to high-throughput sequencing (ChIP-seq) data, and simulations involved 10,000 NFκB molecules (and no polymerases).

Mapping Between Simulation and Physical Units. Energy scales are naturally expressed in units of k_BT (1 k_BT is ∼0.6 kcal/mol). For naked DNA (Figs. 1–3; Figs. S5 and S6), each monomer has a diameter of 2.5 nm, which is the hydration thickness, and represents 7.35 bp of B-DNA (with 0.34-nm axial distance between base pairs), the persistence length is 20 (in simulation units), which is equivalent to 50 nm, and all protein sizes are measured with respect to a unit of length of 2.5 nm. For a 30-nm chromatin fiber (Fig. 5; Fig. S7), compaction is 100 bp/nm (as is common) (5, 6), each monomer represents 3 kbp, and the persistence length (three simulation units, corresponding to 90 nm) gives a fairly flexible fiber. [Estimates of the persistence length of chromatin range from 40 to 200 nm (5, 7)]. For a 20-nm fiber, we use a compaction of 100 bp/nm, so a bead in Fig. 4 corresponds to 2 kbp. Concentrations of proteins and DNA are mapped to real units using the above length mapping and are given in volume fraction and micromolar units (calculated using the total number of simulated proteins and size of cube).

Timescales in simulations can be mapped to physical ones by matching the Brownian time for a DNA/chromatin bead (for example), which equals σ^2/D, to its physical counterpart, obtained by estimating the diffusion coefficient via Stokes' formula

$$D = \frac{k_B T}{3\pi\nu\sigma},$$

where we use a viscosity ν of 1 or 10 cP, which is appropriate for an aqueous solvent or nucleoplasm, respectively. Then, one simulation unit corresponds to 70 (Figs. 1 and 2; Figs. S2, S3, and S5) or 35 ns (Fig. 3) and to 0.36 (Fig. 4) or 1.22 ms (Fig. 5; Fig. S7).

Choice of Initial Configurations in BD Runs. As in all MD simulations, initial conditions play a critical role. When modeling only one polymer, we start with two different initial conditions, which lead to quantitatively and qualitatively similar results. In the majority of simulations, we start with DNA (or chromatin) as a self-avoiding walk and the proteins randomly positioned and unbound. To this end, DNA (or chromatin) was initialized as a random walk and then equilibrated for ∼100,000 steps (0.01 simulation units) with a soft repulsive potential, to generate a self-avoiding configuration. The proteins are initially subject to similar soft repulsive interactions (with both other proteins and DNA). In selected cases, and to compare with patterns found in Fig. 1 and Fig. S5, we start with proteins prebound to specific beads in the polymer through fictitious springs (taking care not to start with clusters). The equilibration then proceeds as before, by subjecting all beads to a soft repulsive potential to eliminate overlaps. After equilibration, potentials are changed to the ones described in Eqs. **S1–S6**, and the fictitious springs are eliminated.

When modeling four human chromosomes, the initial condition has a large impact because the equilibration time of a whole chromosome is longer than currently accessible simulation times (6). Therefore, we use three different initial conditions: (*i*) with intermingled polymers behaving initially as self-avoiding walks (generated as above); (*ii*) with segregated mitotic-like configurations as described by Rosa and Everaers (6); and (*iii*) with segregated self-avoiding chains each occupying a quarter of the simulation box. The results in Fig. 5 and Fig. S7 start with case iii. In all cases clustering proceeds to a similar extent within the time simulated. NFκB is modeled as a sphere with a diameter 30% of that of a 30-nm monomer in the chromatin fiber (i.e., 9 nm); this is chosen to optimize simulation time (because a smaller size leads to similar results but longer simulation times due to the smaller time step required for convergence). The interaction between NFκB and any chromatin bead is set to be either purely repulsive or attractive (Table S2).

Comparison with Chromatin Interaction Analysis with Paired-End Tag Sequencing Data. Contacts seen in simulations involving chromosomes 5, 8, 14, and 17 were compared with those detected by chromatin interaction analysis with paired-end tag sequencing (ChIA-PET) applied with an antibody targeting the active form of RNA polymerase II to HUVECs 30 min after TNFα stimulation (3). One gene (i.e., *SAMD4A*) and two sets of 18 genes on the four chromosomes were selected for detailed analysis. One set included the 18 most up-regulated (determined by comparison of intronic RNA levels determined using RNA-seq and total RNA isolated 0 and 30 min after stimulation); they were *ESM1*, *CCL2*, *NFKBIA*, *RIPK2*, *SDCBP*, *DUSP1*, *IRF1*, *PLK2*, *KCNJ2*, *EIF1*, *STARD4*, *TRIB1*, *SAV1*, *HMGCS1*, *CTHRC1*, *TNFAIP1*, *TNFAIP2*, and *SPRY4*. ChIP-seq confirms that all these up-regulated genes bind the polymerase and p65 (3). The other (constitutively active) set was neither up- nor down-regulated; this set contained *SCFD1*, *RALGAPA1*, *ACTR10*, *SYNJ2BP-COX16*, *ALKBH1*, *MEG3*, *JPH4*, *SPTSSA*, *SNX6*, *MIPOL1*, *SEC23A*, *FKBP3*, *MDGA2*, *TRIM9*, *DACT1*, *RTN1*, *GPHB5*, and *ZDHHC22*. ChIP-seq confirms that all these constitutively active genes bind the polymerase but not p65 (3).

The ChIA-PET data consist of pairs of mapped tags in which each tag in the pair can have up to two mismatches. As tags are short, many map to more than one point in the genome. For Table S1, we first select tags mapping uniquely to the four chromosomes, *SAMD4A*, and the sets of 18 up- or down-regulated genes. We then identify the partner tag (which might map uniquely or nonuniquely to the genome) as a contact. After binning into 3-kbp segments, contacts are compared with those seen in simulations.

1. Plimpton S (1995) Fast parallel algorithms for short-range molecular dynamics. *J Comput Phys* 117(1):1–19.
2. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph* 14(1):33–38.
3. Papantonis A, et al. (2012) TNFα signals through specialized factories where responsive coding and miRNA genes are transcribed. *EMBO J* 31(23):4404–4414.
4. Feng X, Grossman R, Stein L (2011) PeakRanger: A cloud-enabled peak caller for ChIP-seq data. *BMC Bioinformatics* 12:139.

5. Cook PR, Marenduzzo D (2009) Entropic organization of interphase chromosomes. *J Cell Biol* 186(6):825–834.
6. Rosa A, Everaers R (2008) Structure and dynamics of interphase chromosomes. *PLOS Comput Biol* 4(8):e1000153.
7. Langowski J (2006) Polymer chain models of DNA and chromatin. *Eur Phys J E Soft Matter* 19(3):241–249.

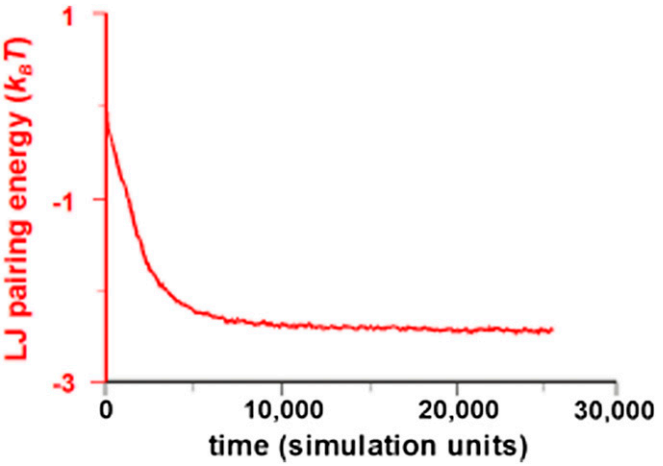


Fig. S1. Changes in pairing energy (sum of Lennard-Jones interactions, either repulsive or attractive) at different times during the simulation in Fig. 1. The decrease in pairing energy per bead (DNA or protein) increases concurrently with the fraction of proteins in clusters (Fig. 1*E*) and reflects the formation of multiple contacts once a protein binds. [For every contact made between a protein and DNA monomer, the energy change is ~4.1 k_BT; therefore, the final energy indicates that (on average) each protein makes around seven to eight contacts (typical configurations show contacts with four DNA segments and with two beads per segment).]

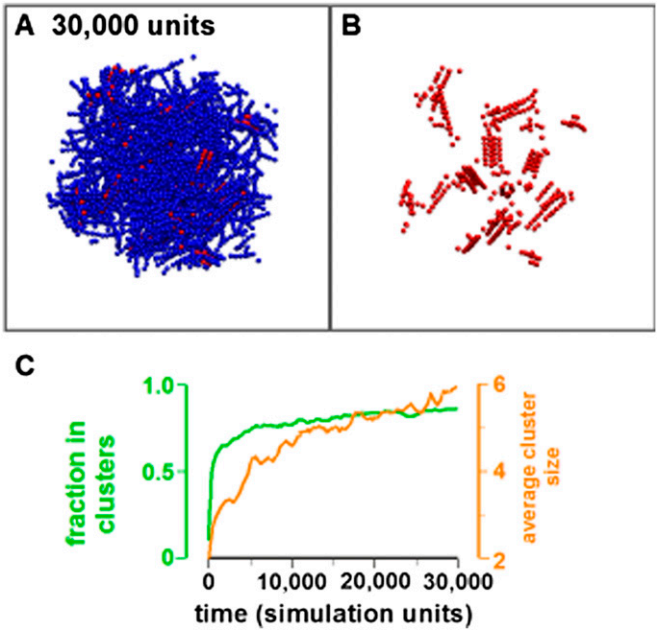


Fig. S2. Clustering also occurs under confined conditions. MD simulations involving one string of blue beads (2.5 nm diameter) representing naked DNA (36.76 kbp; persistence length, 50 nm; volume fraction, 9.7%) interacting with 400 DNA-binding proteins (2.5-nm red spheres; volume fraction, 0.8%) in a cube (75 × 75 × 75 nm). The interaction energy and range (protein bead center to DNA bead center) were 4.12 k_BT and 3.25 nm, respectively. All conditions were as for Fig. 1, except for the higher volume fraction and smaller cube. (*A* and *B*) Two views (with/without DNA) of one structure after 30,000 simulation units (with the same mapping to physical units as in Fig. 1). (*C*) Fraction of proteins in clusters (threshold for clustering 3.5 nm) and average number of proteins in clusters at various times.

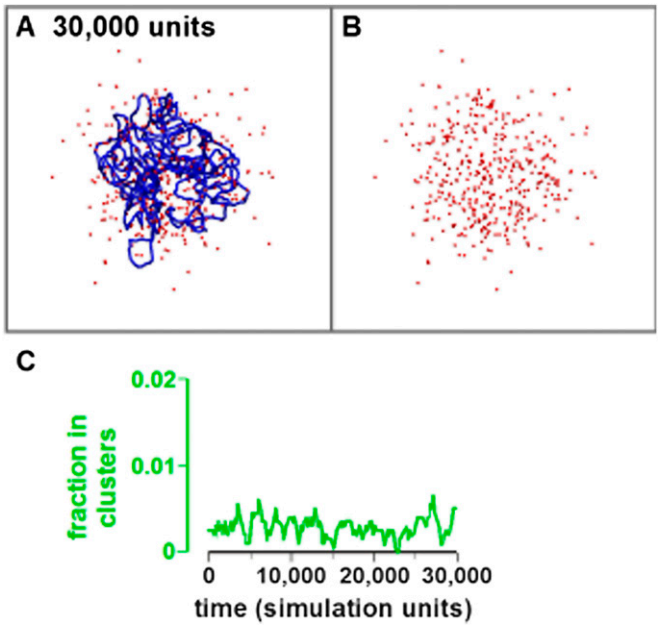


Fig. S3. Clustering requires an attraction (for comparison with Fig. 1). MD simulations involving one string of blue beads (2.5 nm diameter) representing naked DNA (36.76 kbp; persistence length, 50 nm; volume fraction, 0.26%) and 400 DNA-binding proteins (2.5-nm red spheres; volume fraction, 0.02%) in a cube (250 × 250 × 250 nm). All conditions were as for Fig. 1, except that there was no interaction between proteins and DNA. (*A* and *B*) Two views (with/without DNA) of one structure after 30,000 simulation units (same mapping to physical units as in Fig. 1). (*C*) Fraction of proteins in clusters (threshold for clustering 3.5 nm).

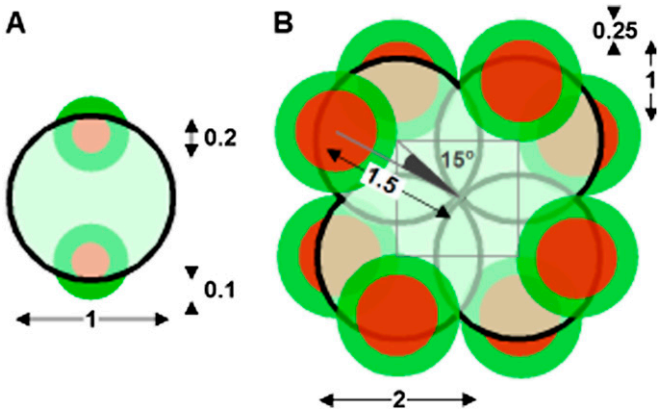


Fig. S4. The structure of patchy proteins (treated as a single entity in simulations). (*A*) Bivalent protein. The large sphere (light green) represents the whole protein (diameter 1σ); it contains two others (red; diameters 0.2σ) touching its north and south poles that have attractive fields (dark green) extending 0.1σ away from their surfaces. (*B*) Nucleosome core. Four large spheres of diameter 2σ (~5 nm) are arranged in a square with centers 2σ apart along the diagonal (so they interpenetrate slightly). Two smaller spheres (red; diameter 1σ) are partially contained within each larger sphere, and an attractive field (dark green) extends 0.25σ away from their surfaces. The centers of the small spheres are 1.5σ from the center of the square, shifted 0.2σ so that there is one small sphere above the plane of the square and one below. The line between the center of each small sphere and the center of the square is at 15° to the line between the center of the large sphere and the center of the square, with small spheres above and below the plane of the square shifted in different directions. As a result, eight binding sites protrude from the circumference of the roughly disk-shaped structure. This structure could, in principle, allow 1.5–2 turns of naked DNA (25–30 beads) to wrap around the circumference of the roughly disk-shaped structure. Although the eight binding sites are chiral, the DNA can wrap around them in either direction (unlike the case in the nucleosome core).

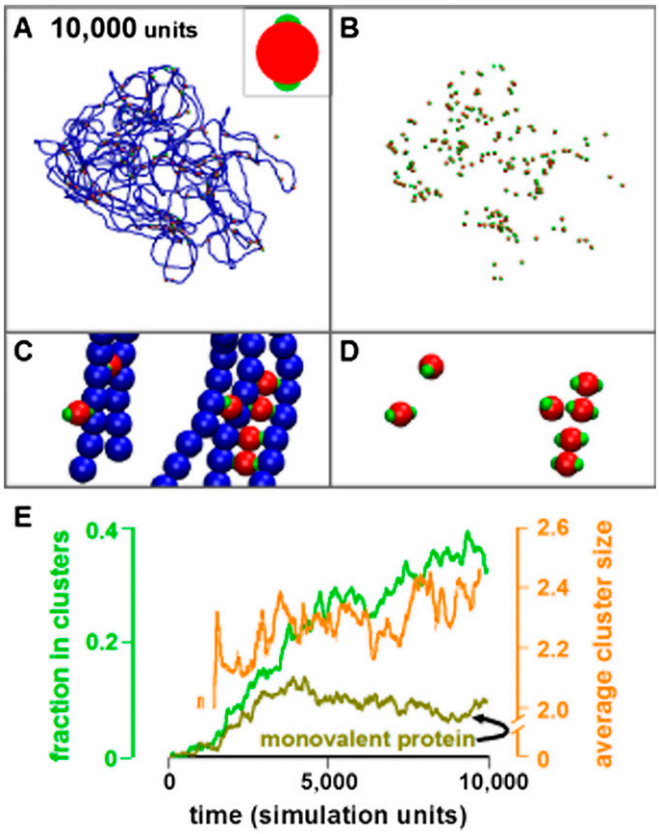


Fig. S5. Bivalent proteins cluster on binding to naked DNA. MD simulations involving one string of blue beads (2.5 nm diameter) representing naked DNA (36.75 kbp; persistence length, 50 nm; volume fraction, 0.26%) interacting (14.69 k_BT; range, 1.71 nm from center of protein bead to DNA bead) with 200 DNA-binding proteins that contain sticky patches at their north and south poles (red spheres of 2.5 nm diameter with green attractive regions at the poles; volume fraction, 0.01%) in a cube (250 × 250 × 250 nm). (*A*) A view of the structure after 10,000 simulation units; most proteins have bound (*Inset*: structure of one protein with its two sticky patches). (*B*) Same view as in *A*, without DNA; some proteins cluster, despite the lack of any explicit protein:protein interaction. (*C* and *D*) High-power view with and without DNA showing clustering into a row on the right. (*E*) Both the fraction of beads in clusters and average cluster size increase with time and reach a steady state, with some fluctuations (two proteins are in one cluster if center-to-center distance is <5 nm). The mustard line shows the fraction in clusters for proteins with only one sticky patch.

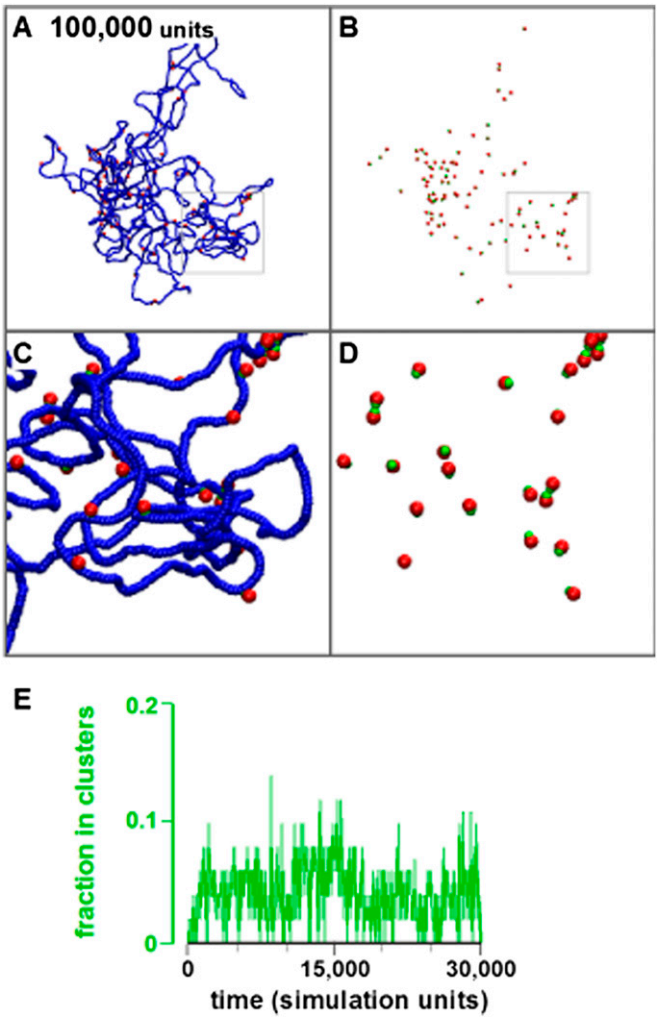


Fig. S6. Spheres with only one binding patch do not cluster appreciably. MD simulations involving one string of blue beads (2.5 nm diameter) representing naked DNA (36.76 kbp; persistence length, 50 nm; volume fraction, 0.26%) interacting with 100 proteins (volume fraction, 0.14%) in a cube (250 × 250 × 250 nm). Each protein (diameter 7.5 nm) has a single DNA-binding site (a patch of 2.5 nm diameter whose center is located 2.5 nm from the center of the protein). The interaction energy and range (DNA center to protein center) were 9.05 k_BT and 3.125 nm, respectively. (*A* and *B*) Two views (with/without DNA) of one structure after 100,000 simulation units; little clustering is observed. Even so, a few clusters are seen; these invariably involve proteins that bind to two DNA monomers (as the binding zone can accommodate binding to two monomers). (*C* and *D*) Two magnified views of the boxed region in *A* and *B*. (*E*) The fraction of beads in clusters changes little with time (two bound proteins are in the same cluster if center-to-center distance is <9 nm); this indicates that bridging is required for clustering.

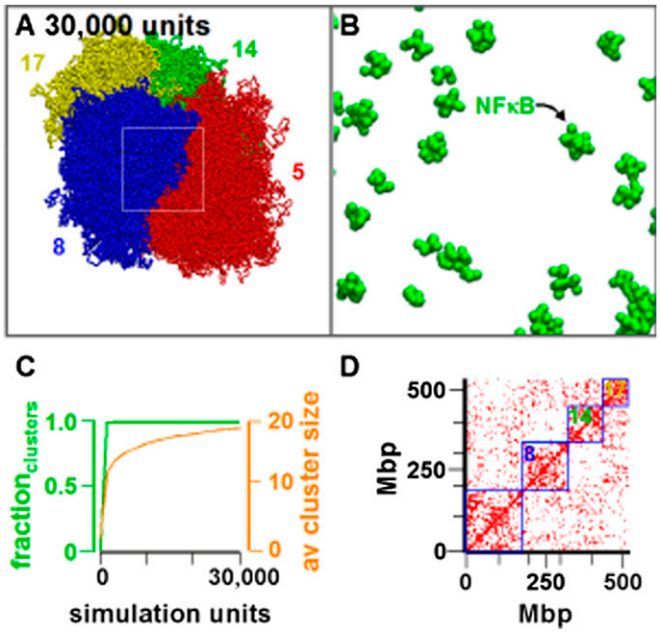


Fig. S7. Clustering of NFκB after binding to human chromosomes 5, 8, 14, and 17 with varying affinity. MD simulations involving four strings of beads (diameter 30 nm) representing human chromosomes 5 (red), 8 (red), 14 (green), and 17 (yellow) modeled as polymers of appropriate length (persistence length, 90 nm; volume fraction, 9%) in a cube ($3 \times 3 \times 3$ μm). The cube also contained 10,000 NFκB molecules (green, modeled as 9-nm spheres) that bind to the appropriate sites on the four chromosomes with one of six different affinities (determined using peak heights in data obtained by ChIP-seq using HUVECs 30 min after stimulation with TNFα). Affinities were 0 (nonbinding), 9.56, 10.82, 11.54, 12.53, or 13.52 k_BT, and the interaction range (between the centers of DNA and protein) was 36 nm. The initial configuration involved four separate and randomly arranged chromosomes in each of the four quarters of a cube. (*A*) A snapshot taken after 30,000 simulation units (equivalent to ~0.5 min of real time, assuming a viscosity of 10 cP, which is appropriate for the nucleoplasm). (*B*) Region in the *Inset* in *A* without chromosomes to highlight protein clustering. (*C*) Both the fraction of beads in clusters and average cluster size increase with time (two proteins form one cluster if center-to-center distance is <36 nm). (*D*) 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 that form more intra- than interchromosomal contacts (blue boxes).

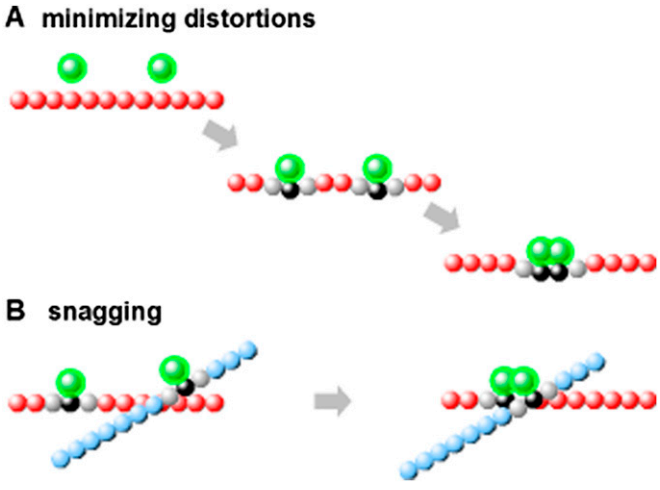


Fig. S8. Cartoon illustrating two effects (the first thermodynamic, the second dynamic) that contribute little to clustering. The cartoons illustrate part of a long polymer and some nearby DNA-binding proteins (green spheres surrounded by attractive zones). When proteins bind, the mobility of black monomers is highly restricted (reducing their entropy); the gray flanking monomers also lose some entropy, and more distant ones progressively less (not indicated). (*A*) Binding of a protein (without bridging) still restricts the mobility of bound monomers, so one might expect clustering of bound proteins to lead to more stable structures (the one on the right has two fewer gray monomers than the one in the middle); however, if this effect is involved, it can only make a minor contribution, as monomeric proteins cluster poorly (Fig. S5). (*B*) A protein bound to one segment is likely to snag against a similar protein bound to another segment as the two segments slide over each other; again, this effect can only make a minor contribution as monomeric proteins cluster poorly (Fig. S5).

Table S1. Comparing contacts seen by simulation and ChIA-PET

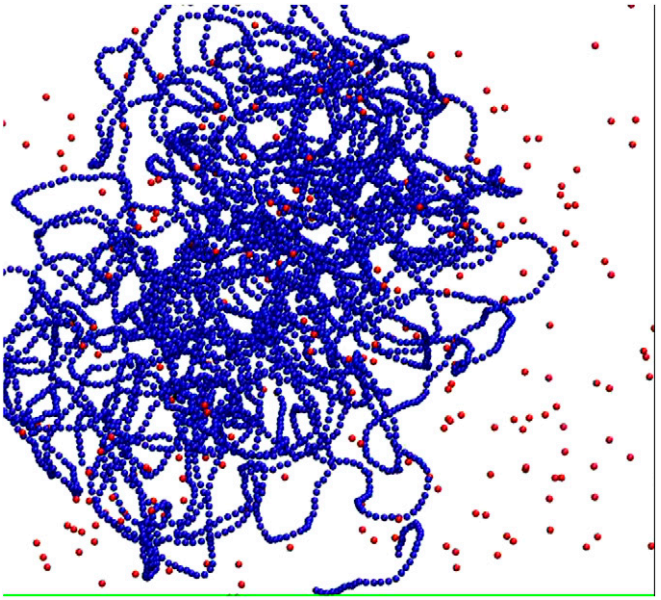
DNA region	Average contacts per bead (3 kbp)		Contacts binding pol II/NFκB (%)	
	Simulation	ChIA-PET	Simulation	ChIA-PET
All four chromosomes	0.244 ± 0.001	0.838 ± 0.001	3.35 ± 0.04	3.35 ± 0.04
SAMD4A	37.9 ± 0.7	22.7 ± 0.5	39.8 ± 1.2	8.7 ± 0.7
Eighteen up-regulated genes	39.5 ± 0.6	45.9 ± 0.8	30.7 ± 0.9	7.5 ± 0.5
Eighteen constitutively active genes	21.8 ± 0.2	0.3 ± 0.02	2 ± 0.1	6.1 ± 1.6

Contacts made by every 3-kbp region of chromosomes 5, 8, 14, and 17 were obtained from simulations (Fig. 5*B*; contact defined as two monomers lying within 90 nm) or ChIA-PET (*SI Methods*); they are roughly similar. Row 1, 3-kbp regions that typically do not bind the polymerase or NFκB make few contacts with other regions; rows 2 and 3, regions binding the two proteins contact many other regions binding one/both proteins; row 4, regions binding just the polymerase make fewer contacts. Errors (±SD) computed assuming Poisson statistics.

Table S2. Parameters for Lennard-Jones (LJ) interactions as in Eqs. S1 and S6

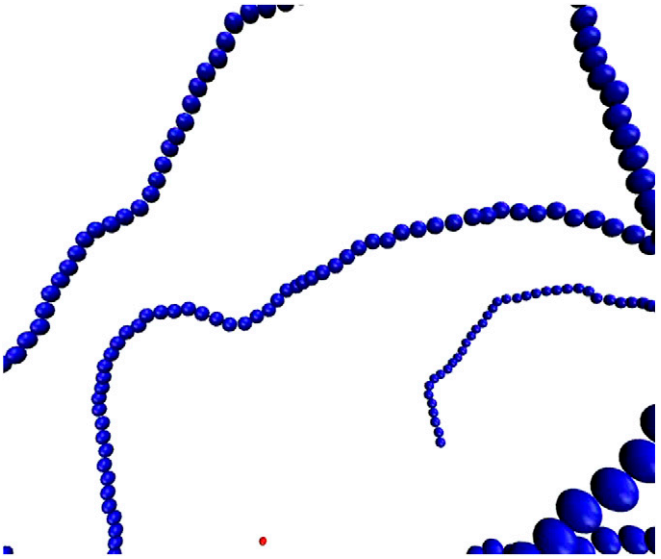
a	*B*	ε_{ab}	σ_{ab}	r_{thr}
LJ parameters for Fig. 1 (1 = protein; 2 = DNA)				
1	1	1.0	1.0	1.122461
1	2	12.0	1.0	1.3
2	2	1.0	1.0	1.122461
LJ parameters for Fig. 2 (1 = protein core; 2 = protein patches; 3 = DNA)				
1	1	1.0	1.0	1.122461
1	2	N/A	N/A	N/A
1	3	1.0	1.0	1.122461
2	2	1.0	0.178	0.2
2	3	50.0	0.535	0.685
3	3	1.0	1.0	1.122461
LJ parameters for Fig. 3 (1 = protein; 2 = DNA)				
1	1	1.0	5.0	5.61231
1	2	3.0	3.0	4.0
2	2	1.0	1.0	1.122461
1	1	1.0	1.0	1.122461
1	2	1.0	1.5	1.683692
LJ parameters for Fig. 4 (1= histone patches; 2= histone cores; 3 = DNA)				
1	3	8.0	1.0	1.4
2	2	1.0	2.0	2.444923
2	3	1.0	1.5	1.683692
3	3	1.0	1.0	1.122461
LJ parameters for Fig. 5 (1 = protein; 2 = DNA)				
1	1	1.0	1.0	1.122461
1	2	12.0	1.0	1.3
2	2	1.0	1.0	1.122461
LJ parameters for Fig. 6 (1 = polymerase II; 2 = p65; 3–6 = DNA)				
1	1	1.0	0.6	0.673477
1	2	1.0	0.45	0.505108
2	2	1.0	0.3	0.336738
1	3 or 5	1.0	0.8	0.897969
1	4 or 6	18.0	0.8	1.44
2	3 or 4	1.0	0.65	0.7296
2	5 or 6	15.0	0.65	1.2
3–6	3–6	1.0	1.0	1.122461

N/A, not applicable.



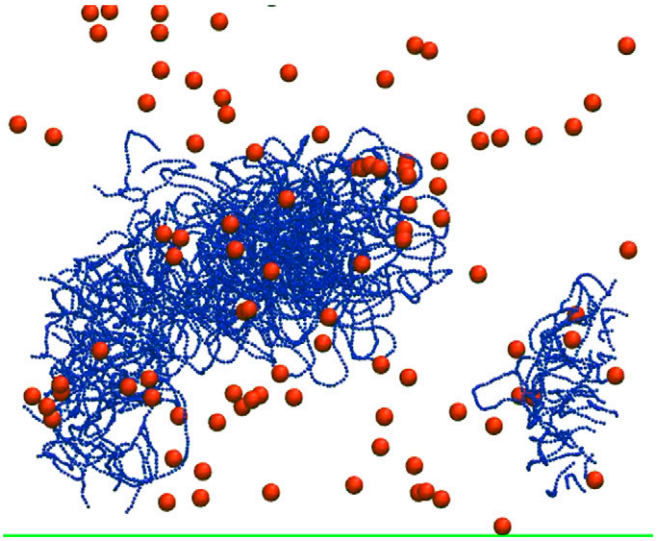
Movie S1. Movie from the simulation in Fig. 1. This movie shows the formation of linear clusters, or rows, as 2.5-nm proteins bind to DNA; clustering occurs in the absence of any explicit attraction between proteins.

Movie S1



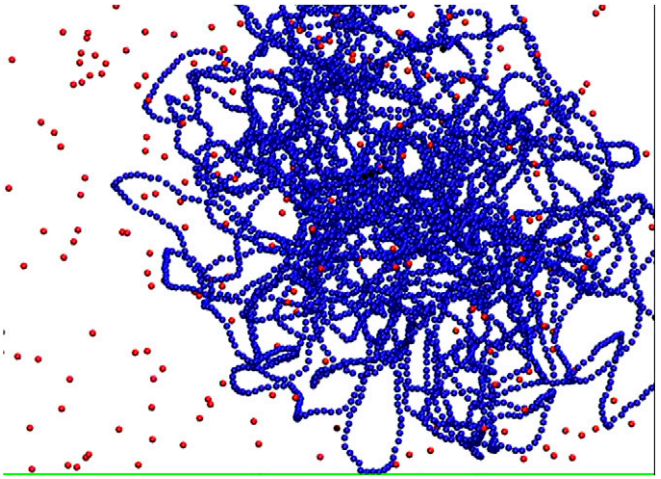
Movie S2. Detail from Movie S1 showing how rows form and the zipper effect. This movie highlights two mechanisms driving row formation. First, two bridges form, which then "zip together". Second, a third protein is "sieved" from the soluble pool to form another bridge between the two other ones (facilitated by the high DNA density and favorable local DNA conformation).

Movie S2



Movie S3. Movie from the simulation in Fig. 2. This movie shows cluster formation for 12.5-nm proteins binding to naked DNA; the shape of the cluster is roughly spherical.

Movie S3



Movie S4. Movie from the simulation in Fig. 4. This movie shows clusters forming as proteins bind to a flexible polymer representing euchromatin; clusters are quasi-spherical.

Movie S4