JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

REVIEW

A Model for all Genomes: The Role of Transcription Factories

Peter R. Cook

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

Edited by J. Karn

A model for all genomes involving one major architectural motif is presented: DNA or chromatin loops are tethered to "factories" through the transcription machinery—a polymerase (active or inactive) or its transcription factors (activators or repressors). These loops appear and disappear as polymerases initiate and terminate (and as factors bind and dissociate), so the structure is ever-changing and self-organizing. This model is parsimonious, detailed (and so easily tested), and incorporates elements found in various other models.

© 2009 Elsevier Ltd. All rights reserved.

The Model

Figure 1 illustrates a model for all genomes that has its roots in observations made by cytologists circa 1900 on meiotic chromosomes and by molecular biologists in the 1970s on isolated bacterial "nucleoids"; active transcription units scattered along a chromosome cluster into "factories", to loop intervening DNA.[1] (For a recent review, see Ref. 2.) Although loops are found in many other models (e.g., Ref. 3), here a promoter distant from a factory is unlikely to be transcribed; it initiates by diffusing to a factory and colliding with one of the polymerases concentrated there. As a result, a loop is transiently tied to a factory through the transcription machinery—a polymerase (active or inactive) or its transcription factors (e.g., activators or repressors such as bacterial CAP and H-NS, or eukaryotic C/EBP and CTCF). Transcription then drives the organization; loops appear (and disappear) as polymerases initiate (and terminate), and factors bind (and dissociate). Reducing the global level of transcription will reduce the total number of molecular ties and so increase average loop

length, and—in cells lacking all transcriptional activity (e.g., starved bacteria, human sperm)—the now-unstructured genome would collapse with neutralizing proteins to occupy the smallest possible volume.[4] Consequently, the structure is dynamic and self-organizing[5,6] with current shape depending on past and present environments. Then, statements about structure are necessarily probabilistic.

One reason for proposing this model is to focus attention on universal architectural features amongst the obvious diversity found in Nature; therefore, many other structural motifs will supplement the ones in Fig. 1. For example, yeast genes diffuse throughout nuclei in minutes, and activity of many is regulated by contacts with nuclear pores; human genes explore much the same absolute volume, but there are proportionately fewer peripheral contacts because nuclear volume is so much greater.[7] In order to retain this focus, non-universal motifs (like such attachments to the periphery) will not be discussed.

Transcription factories play a central role in this model. A "factory" is defined in the *Oxford English Dictionary* as "a building or range of buildings with plant for the manufacture of goods". Factories making nucleic acids were first seen in mammalian cells infected with a single vaccinia virion; on culture in [³H]thymidine, autoradiography revealed one cytoplasmic focus of viral DNA synthesis that grew to contain hundreds of genomes.[8] Such foci were later described as "factories",[9] and virologists were

---

E-mail address: peter.cook@path.ox.ac.uk.

Abbreviations used: EM, electron microscopy; FISH, fluorescence *in situ* hybridization; GFP, green fluorescent protein; 3C, chromosome conformation capture; LCR, locus control region; BrUTP, 5-bromouridine-5′-triphosphate.
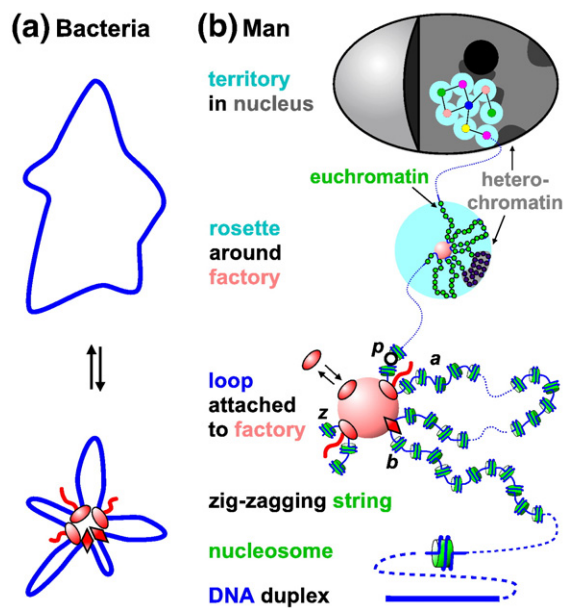
Fig. 1. Model for all genomes. (a) In bacteria, clustering of engaged RNA polymerases (ovals) and transcription factors (diamonds) generates loops; inhibiting transcription disperses nucleoid DNA. Modified from Ref. 6 with permission. (b) In man, DNA is coiled into the nucleosome, and runs of nucleosomes form a zigzagging string looped by attachment to factories through transcription factors and engaged polymerases. [In HeLa, the average contour length of a loop is 86 kb (range, 5–200 kb).] The promoter (*p*) has just initiated, and a fixed polymerase is reeling in its template and will soon transcribe *a*. Components in a factory exchange continually with the soluble pool. About 16 such loops (only a few are shown) form a rosette around a factory (a structure equivalent to the bacterial nucleoid). Distal nucleosomes in long loops tend to be static and acquire a (heterochromatic) histone code that spreads down the fiber; they also aggregate onto the lamina, nucleoli, and chromocenters. A string of 30–180 successive rosettes forms a territory (the general path of DNA is shown). Different factories (circles of different colours) specialize in transcribing different sets of genes. Here, active transcription units that are near neighbours form a rosette (e.g., *a* and *b*), but the structure can be more complex; for example, *z* may be distant from *a* on the genetic map (which would generate a giant loop), perhaps even on a different chromosome. Modified from Ref. 60 with permission from John Wiley & Sons.

using the term before it was applied to sites of replication, transcription, and DNA repair in uninfected cells.[10–12] Each of these factories contained at least two (often more) polymerases and associated plant active on at least two (often more) different templates. (The term "factory" has also been used to describe just one message-producing complex, but it will not be used here in this sense.) The *raison d'être* of all factories is the same: to enhance production by concentrating relevant machines and raw materials in one place. For example, HeLa nuclei contain a 1 μM pool of RNA polymerase II, but essentially all its transcripts are made in factories where the local concentration is ∼1000-fold higher.[6]

A second important feature of this model is the immobilization of engaged RNA polymerases in a factory. Such immobilization runs counter to what we were taught, namely, that active enzymes track like locomotives along their templates. However, there remains little evidence for tracking, which generates an unsolved problem. The resulting transcript is entwined about the helical template once for every 10 bp transcribed (Fig. 2a, left) and must be untwined exactly the right number of times (leaving even one interlock would prevent a eukaryotic transcript escaping to the cytoplasm)— but no satisfactory mechanism for achieving this has been uncovered.[13] This problem simply does not arise if the polymerase is fixed, and reels in its template as it extrudes its (unentangled) transcript (Fig. 2a, right). (Of course, movements are relative, and the polymerase might be fixed to a factory, but both may move together through a cell.) Old experiments first suggested that active polymerases were attached to the nuclear substructure, and so immobilized; most of a loop could be detached with nucleases without removing nascent RNA or transcribed templates.[14] (If polymerases tracked around loops, they should be lost with detached fragments.) Fixed polymerases turn out to be powerful enough molecular motors to act in the required way, and force measurements are now routinely made using single immobilized enzymes.[15] [A much discussed but distinct problem—the removal of the "twin domains" of supercoiling that arise on each side of a polymerase (whether fixed or not)—is solved by topoisomerase action. It is then easy to imagine that topoisomerases are incorporated into the factory on each side of a fixed polymerase.]

Some Theory: Forming Loops and Factories

At least two general mechanisms probably combine to drive looping, and these are outlined in Fig. 2b. One acts transiently through bound transcription factors,[16] and the other for longer through the sheer size of active polymerizing complexes, which can contain a multisubunit enzyme, nascent RNA, and associated proteins such as ribosomes in bacteria or spliceosomes in eukaryotes.[17,18]

In Fig. 1b, neighbouring transcription units a and b are attached to the same factory, and attaching several nearest neighbours creates the neat rosette indicated. However, theory indicates that nearest-neighbour rosettes are only marginally more stable than complex networks in which genes from distant parts of a chromosome fold back to generate larger loops (Fig. 2c). This is consistent with results showing the physical separation between any two human genes in three-dimensional (3D) nuclear space [determined by fluorescence *in situ* hybridization (FISH)] depends on the number of intervening base pairs in a way best fit by mixtures of local and

giant loops of ~ 0.1 and ~ 1 Mbp.[19,20] And just as there are some distant intrachromosomal contacts, there will also be some interchromosomal ones (below).

Such thermodynamic considerations suggest that any piece of transcribed DNA will tend to form loops in crowded cells. Then Nature has its usual choice: to use the resulting structure, or expend energy to forge a new one. It is easy to imagine it chooses the former and goes with the flow. How might a number of such loops come together to create a factory? The textbook view sees active transcription complexes being assembled stepwise from components in a rigid temporal order; however, what is arguably the best-understood functional nuclear structure—the complex involved in nucleotide excision repair—is built differently.[21] It "self-organizes"[5] through random collisions and cooperative binding before potentially useful aggregates are selectively stabilized by "kinetic proofreading". This allows assembly with a specificity above the level available from free-energy differences in intermediates and requires an irreversible step (e.g., involving ATP hydrolysis) that allows the system to have a second go at discriminating between wanted and unwanted structures.[22] By stringing together many such steps, progressively larger complexes can be built. In the case of a transcription factory, once a few bound polymerases/factors come together (as illustrated in Fig. 2b and c), the increased local concentration of binding sites makes it more likely others are recruited. But as more and more loops are added, modeling indicates the resulting high DNA density at the surface soon limits further growth (D. Marenduzzo and P.R.C., unpublished data). So although a factory might contain sub-

assemblies (such as polymerase:mediator complexes) with defined 3D structures, the whole would be pleiomorphic and intrinsically unstable, persisting (like a cytoskeleton) only by exchanging
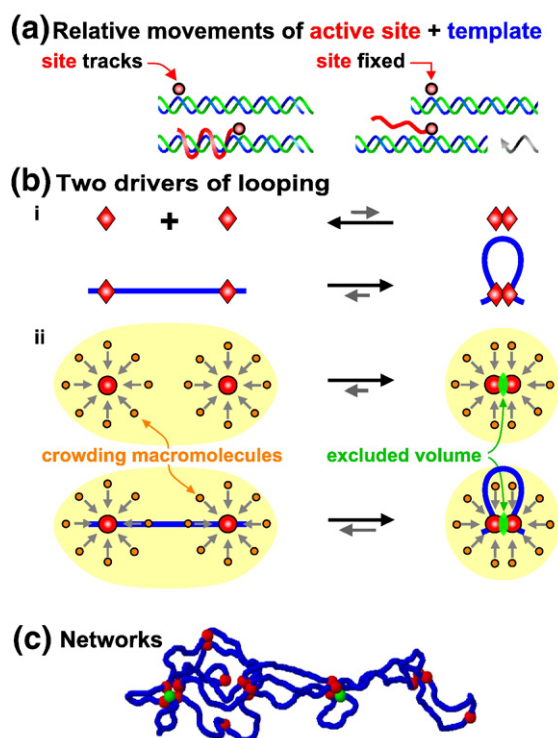

Fig. 2. Theory: immobile polymerases, and loop formation. (a) Relative movements of the active site of a polymerase (circle) and helical template (the transcription bubble, and resulting supercoils are not shown). In the conventional model for transcription (left), the active site in a tracking polymerase moves around and along as it reads a helical strand; the resulting transcript (red) is entwined about the template and must be untwined exactly the right number of times before it can escape. No satisfactory mechanism for doing this has yet been identified, and this represents a major (but rarely discussed) shortcoming of the conventional model. An alternative that sidesteps this problem has the active site in a fixed polymerase (right) reeling in its template (which moves around and along, indicated by the grey arrow), as an unentangled transcript is extruded. This alternative is a central feature of the model for all genomes proposed here. Modified from Ref. 13 with permission. (b) Two drivers of looping. (i). Specific interactions.[16] Top: If two DNA-binding proteins (which could be the same or different molecules) are present at ~ 1 nM and interact with a K_d of 10^{-7} M (values typical of transcription factors), <1% come together (so the equilibrium is to the left). Bottom: On adding a DNA molecule with two binding sites 10 kb apart, protein binding to the template creates a local concentration that drives two-thirds into the complex (inevitably forming a loop); the equilibrium is now to the right. Such loops are unlikely to persist for long, as GFP tagging shows factors typically reside on DNA for <10 s.[7] (ii) Non-specific (entropic) "depletion attraction". Top: In a crowded cell, many small soluble macromolecules (brown) bombard large complexes from all sides (grey arrows). When two complexes come into contact, small macromolecules are sterically excluded from the green volume between the two and so cannot knock the two large complexes apart; as a result, a "depletion attraction" (equivalent to the osmotic pressure exerted by small macromolecules on opposite sides of the two large complexes), keeps the large complexes together. Bottom: When the large spheres (polymerases) are threaded on a string (DNA or chromatin), this depletion attraction is only partially countered by the entropic cost of looping. It has the strength of a few H-bonds and will act for as long as polymerases remain engaged. This can be seconds in bacteria, and minutes in man (longer if polymerases pause, as many do).[54] This "attraction" can act in the absence of forces familiar to biologists (i.e., those involving H-bonds, van der Waals forces, hydrophobic and charge interactions that might underpin the interaction in (i)), but may be supplemented by them. Adapted from Ref. 35. (c) Forming networks. Monte Carlo simulation of 21 beads (terminal beads are green, internal ones red) threaded every 20 kb along a (self-avoiding) 0.4-Mb chromatin fiber (blue). Each bead represents a transcription complex (i.e., 15-nm RNA polymerase II, 20-nm transcript and associated proteins, 24-nm spliceosome). Starting with a linear string, fiber segments are allowed to "diffuse" in a computer while being subjected to a "depletion attraction" between any two beads of 4 k_BT. [Other interactions between bound transcription factors, such as those in b(i), could also generate an equivalent force.] After reaching equilibrium, a typical structure is shown. Most beads are in "factories", but many contacts between non-nearest neighbours are seen. From Ref. 18.

subunits with others in its surroundings—as tagging with the green fluorescent protein (GFP) indicates.[23,24] Experimental evidence for the existence of loops and factories is now summarized.

Prototypic Factories at the Core of Bacterial Nucleoids

Nucleoids are released by treating bacteria with lysozyme, a detergent, and 1 M NaCl; they contain rosettes of naked (supercoiled) DNA attached to clusters of engaged polymerases, as in Fig. 1a.[25] The structure is maintained by transcription, since it unfolds on pretreatment with an inhibitor of the bacterial RNA polymerase, rifampicin, or posttreatment with RNase. Loops in a rosette are topologically distinct, as a single-strand break releases supercoiling only in one loop. There was initially no suggestion equivalent structures existed *in vivo*, presumably because it was assumed that immobilized polymerases could not work, and that tracking enzymes must have aggregated artifactually to generate the structure.

Nevertheless, loops do preexist *in vivo*, as nicking progressively reduces binding of [³H]trimethylpsoralen, a probe for supercoiling (and so looping). But which molecular ties maintain such loops?

Ties in living *Salmonella typhimurium* have been mapped using site-specific recombination and the approach illustrated in Fig. 3a; most attached sequences turn out to be active genes,[26] which suggests engaged polymerases are major ties. Then,

the contour length of loops should roughly equal interpolymerase spacing, and results obtained using different methods (i.e., site-specific recombination, microarrays, and electron microscopy) show both are between 10 and 20 kb.[26,27] For example, transcription complexes are typically spaced this far apart on spread DNA fibers.[28] (The tightly packed polymerases seen in the iconic "Christmas trees" in such "Miller" spreads are the exceptions—the highly transcribed ribosomal DNA (*rrn*) operons.) If active polymerases also cluster into factories, GFP tagging of the polymerase should reveal this; again, it does (Fig. 3b).[29] Many other proteins probably act as ties in special cases, including "structural" proteins such as H-NS, IHF, and FIS (significantly, all are transcription factors bound with polymerases at promoters,[30]), DNA topoisomerases and translocases (gyrase, FtsK, MukBEF), and the actin homolog MreB.[26,31]

Active Transcription Units as Molecular Ties in Eukaryotes

At the end of the 19th century, microscopists observed loops attached to chromomeres in meiotic cells (during stages we now know are transcriptionally active), but the first evidence for looping of eukaryotic genomes during interphase came from the demonstration of supercoiling in human nucleoids.[32] Again, a single-strand break released supercoiling from only one loop, showing the linear chromosome was tied into topologically distinct domains. Significantly, supercoiling in derivative
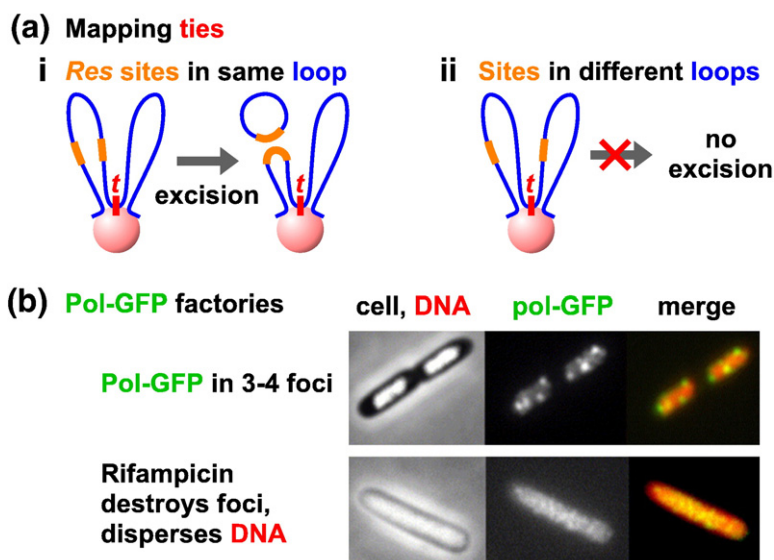

Fig. 3. In bacteria, active transcription units act as ties, and cluster. (a) Mapping ties (*t*) in living cells using site-specific recombination.[26] (i) The γδ resolvase can excise DNA between two *res* sites (brown) in one loop; one half of the dimeric enzyme binds at one site, while the other scans along DNA until it finds a second (to bring the two together). Excision is readily monitored when the excised DNA encodes a selectable marker. (ii) No excision is detected when two sites lie on opposite sides of any barrier/tie that blocks scanning. Barriers/ties can be mapped by inserting *res* sites progressively further and further apart on a chromosome; on inducing resolvase, excision occurs until a barrier lies between the two. Active genes prove to be major barriers. A decisive experiment then confirmed the act of transcription creates a barrier. An inducible promoter driving *lacZ* was inserted between two sites. When uninduced, DNA was excised efficiently; when induced, it was not (as a transcription-based barrier/tie now blocked scanning). Modified from Ref. 35 with permission. (b) Factories. RNA polymerase was tagged with GFP, cells were fixed and stained with 4′,6-diamidino-2-phenylindole (DAPI), and imaged. Under the conditions used, ~70% of polymerases active in a cell are engaged on ~22 *rrn* operons (each with ~70 polymerases), but only three to four GFP foci are seen; many operons must cluster into fewer foci. Adding a transcriptional inhibitor, rifampicin, destroys foci and disperses DNA. Modified from Ref. 29 with permission.

nucleoids was lost progressively as transcriptionally active chicken erythroblasts matured into inert erythrocytes, which again pointed to a critical role for transcription in maintaining loops. Unrestrained supercoils were also found in living insect and human cells.[33,34]

If active RNA polymerases are major ties, endonucleases should detach the bulk of most loops to leave those polymerases, their templates, and nascent RNA (Fig. 4a). Early studies on nucleoids showed templates and nascent RNA remained,[32] but these results could be criticized on the grounds that active transcription complexes had aggregated artifactually in the unphysiological buffers used. However, results were confirmed using cells permeabilized in "physiological" buffers.[6,35] In contrast to results obtained with structures such as "matrices" and "scaffolds" (both prepared using unphysiological conditions), analysis of residual fragments obtained in "physiological" buffers provided no evidence that any particular DNA sequence or motif—other than the ones of interest here—was involved in maintaining most loops. Thus, removing all but 10% of cellular DNA should enrich a sequence attached in every cell in the population 10-fold; however, enrichments were always much less. Moreover, the particular DNA sequences and proteins constituting the molecular ties are unlikely to be conserved, as the various genome sequencing projects have signally failed to uncover any likely candidates.

Chromosome conformation capture (3C) and FISH now convincingly show loops to be transiently tied through active transcription units. 3C and its derivative techniques involve fixation before analysis of which sequences tend to lie next to each other in 3D nuclear space. Contacts made by mouse *Hbb-b1* (encoding β-globin) have
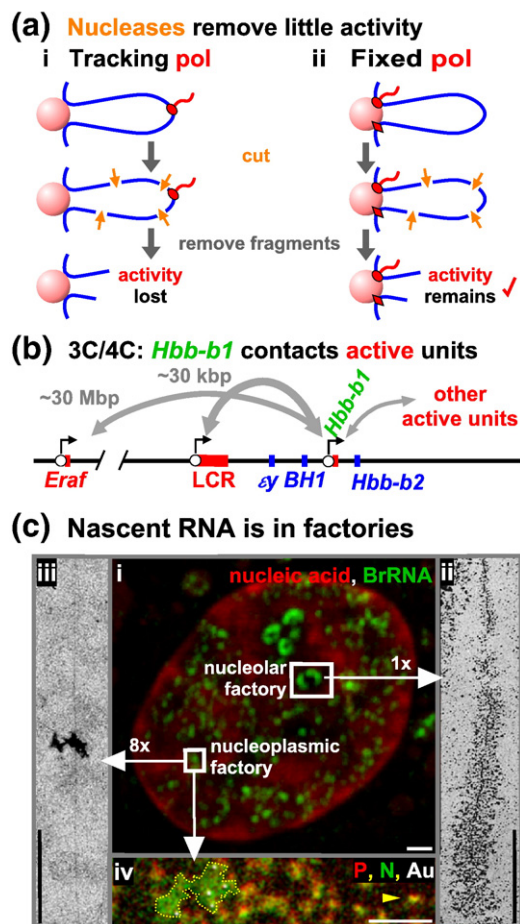

Fig. 4. In mammals, active polymerases act as ties, and cluster. (a) Are active polymerases attached to the substructure? Cells are permeabilized in a "physiological" buffer, treated with nucleases to cut the chromatin fiber (brown arrows), and detached chromatin is removed electrophoretically. (i) If polymerases track, all activity should be lost. (ii) If attached, all activity (plus nascent RNA/templates) should remain; they do.[35] (b) 3C and its derivative (4C) show *Hbb-b1* in mouse fetal liver often contacts its distant LCR and *Eraf*, but not many nearer genes.[38] The thickness of the grey arrows reflects contact frequency (and so looping frequency). (c) Factories in HeLa cells. (i) Cells were permeabilized, nascent transcripts extended by ~40 nucleotides in BrUTP, cells cryosectioned (100 nm), resulting BrRNA was immunolabeled with fluorescein isothiocyanate (green), nucleic acids were counterstained with TOTO-3 (red), and a confocal image was collected. Newly made BrRNA is concentrated in factories in the cytoplasm (made by mitochondrial polymerases), nucleoplasm, and nucleoli. The brightest foci are generally in nucleoli, which specialize in ribosome production. Human rDNA loci are carried on five chromosomes, with each locus encoding ~70 tandem repeats of the 45S rRNA gene. During mitosis, some loci are associated (i.e., "book-marked") with (inactive) RNA polymerase I and its transcription factor, UBF, and these fuse to form nucleoli when transcription begins during telophase; others lacking the two remain inactive and do not fuse.[61] Nascent rRNA within resulting nucleoli is often found in a crescent-shaped "dense fibrillar component" on the surface of an underlying "fibrillar center" (box).[55] From Ref. 13; image courtesy of A. Pombo. The scale bar represents 1 μm. (ii) Stripping off and spreading one of about four crescents typically found in a nucleolar factory yields the iconic EM image of a "Christmas tree" with ~125 tightly packed polymerases. Then, such a factory contains ~500 polymerase I molecules engaged on about four transcription units.[44] The scale bar represents 1 μm. From Ref. 62 with permission of the Society of the European Journal of Endocrinology. (iii) Stripping off and spreading one of about eight active transcription units in a nucleoplasmic factory yields this EM image (with one polymerase engaged on its template). The scale bar represents 1 μm. From Ref. 44. (iv) Image of nucleoplasmic factory in an unstained section obtained using an electron microscope with a special filter. Cells were permeabilized, nascent transcripts extended in BrUTP, and resulting BrRNA was immunolabeled with 5-nm gold particles; after sectioning (70 nm), images of endogenous phosphorus (red) and nitrogen (green) plus immunolabeling gold particles (white) were collected and merged. Five particles mark BrRNA in a nitrogen-rich factory (perimeter indicated). Absolute numbers of N and P atoms within this perimeter can be calculated using nearby nucleosomes as references (arrowhead). The scale bar represents 100 nm. From Ref. 46.

been studied extensively,[36] and so they will be discussed; however, analogous results have been obtained with many other loci.[37,38] When *Hbb-b1* is silent (in brain), only 13% of its contacts are with other active units, but when transcribed (in fetal liver), 80% are now with other active units [including the locus control region (LCR)]. Most are with nearby units, although some are with distant segments of the same chromosome and fewer still with other chromosomes (Fig. 4b). FISH confirms relevant templates and their nascent transcripts tend to be together in factories (sometimes called "active chromosome hubs" in this context).[38] (As even highly expressed *Hbb-b1* is transcribed sporadically, at any moment only some alleles in the population make such contacts.) Factors such as EKLF, GATA-1, and FOG-1 act as additional ties, and these can probably maintain loops for short periods in the absence of transcription because some *Hbb-b1*–LCR interactions persist when transcription is inhibited with 5,6-dichloro-1-β-d-ribofuranosyl-benzimidazole.[39,40] [This inhibitor drives engaged RNA polymerase II off the template,[23] and the persistence of loops contrasts with what is seen with rifampicin in bacteria, which seems to destroy looping mediated by the clustering of active polymerases (above).] "Dam identification" provides further evidence for loops in living cells.[41] As in bacteria, the high DNA concentration at the heart of rosettes would attract other stabilizing ties, including transcription factors such as fly GAGA factor, Su(Hw) insulator, Zw5 and BEAF, chicken CTCF, mammalian SATB1, plus other components such as topoisomerase II, condensins, and cohesins.[42]

Specialized Eukaryotic Factories Containing RNA Polymerases I, II, and III

Imaging also provides excellent evidence for clustering of nascent RNA in factories dedicated to producing transcripts made by either polymerase I, II, or III (Fig. 4c).[43] In various higher eukaryotes (e.g., HeLa, undifferentiated and differentiated mouse ES cells, newt cells with 11-fold larger genomes), careful quantitative analysis shows that a typical polymerase II factory contains approximately eight active polymerases, each engaged on a different unit.[44,45] Nanoscale mapping of phosphorus and nitrogen in or around such factories using a special electron microscope reveals templates and transcripts on the surface of a protein-rich core; these cores are polymorphic, with a diameter of ~87 nm and mass of ~10 MDa [Fig. 4c(iv)].[46] Little RNA is made outside factories, as ~95% label marking nascent RNA is associated with factories (estimated using conditions where essentially all nascent RNA is labeled after allowing engaged polymerases to "run on" by ~40 nucleotides).[43,45] Unfortunately, these factories have not yet been purified, probably because they are both sensitive to current extraction

procedures while being tightly attached to an internal lamin-containing nucleoskeleton.[11,47]

Some polymerase II factories specialize further and transcribe different gene subsets. For example, 3C and FISH show that different transcription units encoding components involved in the globin pathway (e.g., *Hbb-b1*, its LCR, and *Eraf*) on mouse chromosome 7 are often (but not always) together in factories when active (above). Moreover, two mini-chromosomes carrying essentially identical units are transcribed in the same factories, but inserting into one a different promoter (or intron) now targets it to a different factory.[48] It remains to be seen to what extent factories specialize and how many different kinds there might be, but it is already clear co-regulated genes often cluster on a chromosome[49]—consistent with them tending to be co-transcribed in the same specialized factory.

Loops and Gene Regulation

Time-lapse imaging of LacO-tagged loci (in living yeast, fly, and human cells) indicates DNA can diffuse freely throughout a local nuclear volume (diameter 0.5–1 μm), but is then restrained from diffusing further afield.[7] (Movements over longer distances, sometimes involving actin motors, are also seen.[7]) Intuition would then suggest a promoter tethered close to a factory would be more likely than a distant one to collide with that factory and initiate. Simulations confirm this (Fig. 5a),[50] and it is then easy to imagine that "hot" (proximal) and "cold" (distal) promoters will be euchromatic and heterochromatic, respectively, and that regulatory motifs would act by tethering target promoters more or less closely to factories.[51] Figure 5b illustrates the way "enhancers" and "silencers" might work, while "barriers" (not shown) would be active transcription units often attached to a factory—and this would prevent spread of (inactivating) heterochromatin down a fiber. These notions are consistent with old ideas that these motifs loop back to contact target promoters[52] and with recent 3C/FISH results demonstrating such contacts (e.g., between *Hbb-b1* and its LCR; see above). Moreover, canonical enhancers (e.g., the globin LCR), silencers (e.g., yeast tRNA genes), and barriers (e.g., yeast HMR, fly *scs* and *scs'*) all turn out to be the active transcription units required for this kind of explanation (reviewed in Ref. 51). And, significantly, the particular name we give to such motifs (whether enhancer, silencer, etc.) critically depends on the "context", which is consistent with the model in Fig. 5b.[51]

How might loop structure change when, for example, mouse ES cells differentiate into parietal endoderm? Quantitative analysis shows the number of active polymerases, factory number, and nuclear volume all fall, factory diameter and density remain constant, and the amount of heterochromatin increases.[45] This means loop volume remains the same as loops become longer, and the system is able
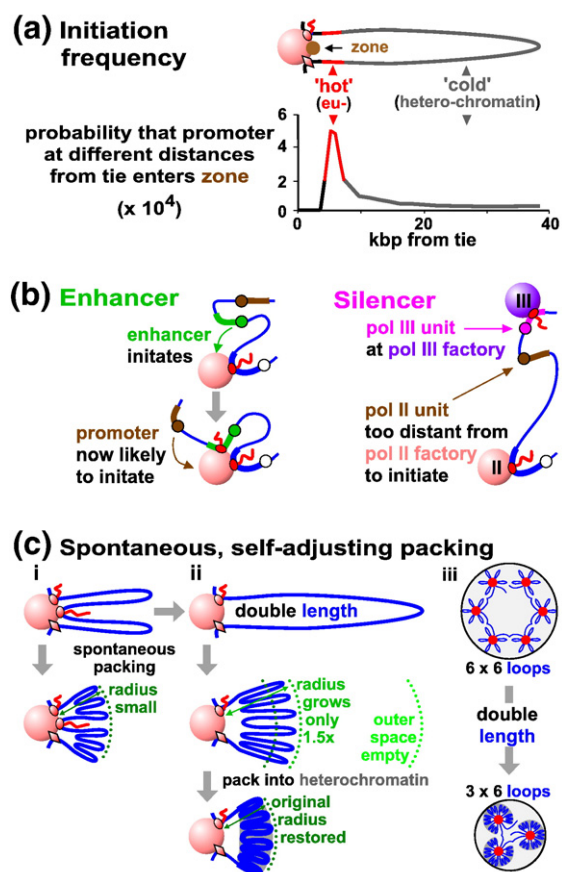
Fig. 5. Models for gene regulation. (a) Initiation frequency. Monte Carlo simulations indicate a "hot" promoter in a proximal segment (red) in a typical human loop is more likely to collide with a polymerase in the brown zone on the surface of a factory (and so initiate) than a "cold" one in a more distant segment (grey). Proximal and distal segments would then be euchromatic and heterochromatic, respectively. Adapted from Ref. 50, with permission. (b) Enhancers and silencers (canonical examples are transcription units[51]). Enhancer: transcription of the proximal enhancer (green) brings its target gene (brown) into a "hot" segment near the factory, increasing the chances the target will initiate. Silencer: a polymerase III unit (mauve) silences the polymerase II unit (brown) by tethering the brown unit near a polymerase III factory (purple) but far from the appropriate (polymerase II) factory (pink). (c) Self-regulatory changes occurring during differentiation (assuming numbers of active polymerases/genes—and so ties—halve). (i) Monte Carlo simulations show a loop spontaneously packs into a shell immediately around a factory.[50] (ii) On differentiation, loop length doubles (as the number of ties halves), but the radius of the occupied volume increases only ~1.5-fold; when the distal DNA in the now longer loop packs into heterochromatin, the original radius is restored. (iii) At the global level, halving the number of active polymerases per nucleus doubles loop length and halves the number of loops plus rosettes, when factory diameter and density remain constant (here, six rosettes each with six loops give three rosettes with loops of twice the length); nuclear volume also halves, as loop volume, factory density, and the activity of individual factories all remain the same. (Adapted from Ref. 45).

to adjust packing accordingly. A self-adjusting mechanism for achieving this, which is consistent with both experimental and theoretical results, is illustrated in Fig. 5c.

FAQs

Are all loops tied by active polymerases? No. Many are tied by transcription factors and other components of the transcription machinery (Fig. 1). For example, some *Hbb-b1*–LCR interactions persist when transcription is inhibited.[39,40] However, it remains unclear exactly what fraction of loops are tied through polymerases, factors, and other components. Thus, in bacteria, >50% ties detected using the resolvase assay (Fig. 3a) involve active transcription units[26]—consistent with engaged polymerases being major ties—while roughly half the mini-chromosomes in transfected monkey cells are attached through the body of a transcription unit and half through non-transcribed promoters, suggesting attachments are split roughly equally between active polymerases and transcription factors or inactive polymerases.[53] In higher eukaryotes, it also seems likely that many loops will be stably tied through poised or paused polymerases.[54] And, because the focus here is on universal ties, a plethora of additional ones will create loops in specific cases. (Note such ties often involve transcription factors.[42])

Do preformed inactive factories exist? Although there are always exceptions in biology, there still seems to be no good evidence for their existence.

How can closely spaced, divergent, genes be transcribed? One (unbent) template cannot travel in opposite directions at one time, so two genes on it cannot be transcribed simultaneously. This is in accord with what is found. The term "transcriptional interference" is used to describe the underlying phenomenon,[51] and the zero probability seen in the curve in Fig. 5a provides a simple explanation for it. But the template can be reeled in first in one direction, and then in the opposite one, so our two genes can be transcribed successively. The two could also be transcribed simultaneously in the special case where the template is bent in a "U" around a factory; two appropriately facing polymerases on the surface could then initiate simultaneously and elongate together as each reels in one arm of the "U".

Can a gene be transcribed by more than one polymerase? Yes. There is no theoretical reason why one template cannot pass through two or more polymerases on the surface of one factory, and this is how ribosomal DNA operons seem to be transcribed—as a promoter is extruded from one polymerase, it immediately initiates at another in the same factory [with the resulting transcripts giving a crescent in Fig. 4c(i)].[55] (Then there is need for only one topoisomerase per polymerase set to remove resulting positive and negative supercoils.) In contrast to this special case, it seems that most protein-coding genes in bacteria and yeast (even highly active ones) initiate so rarely it is unlikely that two or more

polymerases will be found productively engaged on one gene at any moment.[27,28]

Could a factory simply represent a cluster of polymerases engaged on one highly active gene? The evidence is against this, at least in the typical case (again, there are probably exceptions!). For example, each bacterial focus/factory in Fig. 3b contains roughly six rDNA cistrons, and each nucleolar factory in Fig. 4c(i) about four. And whenever 3C and FISH show two sequences lying together, both generally turn out to be transcriptionally active—for example, *Hbb-b1* plus its LCR, or *Hbb-b1* plus any other gene on the same or different chromosome.[37,38] Moreover, quantitative analyses [using different labeling procedures and both light microscopy and electron microscopy (EM)] show there are typically approximately eightfold more engaged polymerases and their transcripts than nucleoplasmic factories in mammals.[43,45] As (in bacteria to man) an active transcription unit seems to be rarely associated with >1 engaged polymerase [as in Fig. 4c(iii)],[27,28] this means there are typically approximately eight different transcription units per factory.

How can this model be reconciled with images of polymerases apparently tracking around lampbrush loops? Lampbrush chromosomes can be isolated from oocytes of many animals (but not yet from mammals).[56] During the first meiotic division, duplicated homologs pair, and long loops can be seen extending micrometers away from axial chromomeres. Unlike most chromosomes that are characterized by transcriptional inactivity, these ones are highly active—and nascent transcripts can be seen attached to loops. However, many nascent transcripts are also associated with chromomeres. As such loops only become visible on dispersing chromatin in unphysiological buffers, and as none are seen in sections of whole oocytes (where chromatin appears as a granular aggregate), these loops could be created by stripping active transcription units off factories in much the same way the "Christmas tree" in Fig. 4c(ii) was generated from one compact nucleolar crescent. Note the transcriptional inhibitor, actinomycin D, prevents lampbrush loops from forming when sperm heads (which contain unlooped DNA) are injected into frog oocytes.[57]

How do mitotic chromosomes form? Any successful model for genome organization should be able to explain how an interphase (roughly spherical) chromosome territory condenses into a (cylindrical) mitotic chromosome. As the contour length of loops and the basic shape of a territory remain unchanged on entry into mitosis, rosettes could pack on top of each other as they condense.[1] [Entropic forces may assist such condensation (D. Marenduzzo and P.R.C., manuscript in preparation).]

Do "globin" factories only transcribe erythroid-specific genes? Probably not, as erythroid-specific genes may well visit "non-globin" factories and initiate there (but with a low probability).

Are there fewer factories in primary mouse fetal liver cells compared to established HeLa/ES cells? Accurate counts can only be obtained when most factories are detected, and (to date) conditions for ensuring this have only been developed for established lines; they indicate the density of nucleoplasmic factories is constant despite variations in nuclear volume (numbers increase as volume increases; Fig. 5c). Then, numbers seen in primary cells[37] are minimum values. [The necessary conditions are best established after permeabilizing cells, and allowing engaged polymerases to run on in BrUTP (5-bromouridine-5′-triphosphate) so that they synthesize twice the length of (now labelled) transcript. If all factories are immunodetected, the same number should be seen (each with increased labeling); if some lie below the detection threshold, increased incorporation will bring more above the threshold.]

Can genes on different chromosomes be transcribed in one factory? Yes; results of 4C (Fig. 4b) and simulations such as those in Fig. 2c show this to be so.[38,58] However, they also confirm the intuition that a gene is more likely to contact its nearest neighbours on the same chromosome compared to those on other chromosomes. Nevertheless, it may be that transcription of similar genes in one specialized factory underlies pairing of homologous chromosomes (e.g., during meiosis, transvection in flies, and X-chromosome inactivation in mammals).[59]

Will two daughter cells in a clonal population possess the same constellation of loops and factories? Probably not, simply because binding to factories occurs stochastically, and there are so many binding sequences/factories. Even so, two roughly similar constellations in daughters would nevertheless determine roughly similar expression patterns—just as two different constellations would probably encode very different patterns. And unravelling the underlying fuzzy logic will represent a real challenge!

Conclusions

Life forms concentrate molecules to promote interaction, with particular cellular compartments specializing in particular processes, such as mitochondria in energy production. This model extends this principle to transcription. It also has many other advantages. It is general (applying to all genomes), parsimonious (involving only one architectural motif—a DNA–chromatin loop transiently tethered to a transcription factory through the transcription machinery), comprehensive (e.g., it explains how the organization changes during differentiation), inclusive (e.g., it incorporates elements found in other models such as enhancers being tethered to their target promoters by transcription factors), and detailed (e.g., molecules acting as critical ties are specified). It is in tune with recent discoveries that transcription factors and polymerases bound in/around promoters (many apparently non-productively) play important roles in organizing the structure.[42,54] It is also predictive (e.g., it provides simple and unifying explanations of how mysterious

elements such as enhancers, barriers, and silencers might work, and how chromosomes might pair). Most importantly, it is easily tested. For example, the combination of various techniques, including chromatin immunoprecipitation, 3C, and nuclear run ons, with "deep" DNA sequencing, will soon provide us with the relative frequency with which every base in a genome interacts with a polymerase, various transcription factors, and every other base (through looping). We should then soon know which aspects of this model prove to be correct.

References

1. Cook, P. R. (1995). A chromomeric model for nuclear and chromosome structure. *J. Cell Sci.* **108**, 2927–2935.
2. Sutherland, H. & Bickmore, W. A. (2009). Transcription factories: gene expression in unions? *Nat. Rev. Genet.* **10**, 457–466.
3. Pienta, K. J. & Coffey, D. S. (1984). A structural analysis of the role of the nuclear matrix and DNA loops in the organization of the nucleus and chromosome. *J. Cell Sci. Suppl.* **1**, 123–135.
4. Frenkiel-Krispin, D. & Minsky, A. (2006). Nucleoid organization and the maintenance of DNA integrity in *E. coli*, *B. subtilis* and *D. radiodurans*. *J. Struct. Biol.* **156**, 311–319.
5. Misteli, T. (2001). The concept of self-organization in cellular architecture. *J. Cell Biol.* **155**, 181–185.
6. Cook, P. R. (2002). Predicting three-dimensional genome structure from transcriptional activity. *Nat. Genet.* **32**, 347–352.
7. Wachsmuth, M., Caudron-Herger, M. & Rippe, K. (2008). Genome organization: balancing stability and plasticity. *Biochim. Biophys. Acta*, **783**, 2061–2079.
8. Cairns, J (1960). The initiation of vaccinia infection. *Virology*, **11**, 603–623.
9. Joklik, W. K. (1968). The poxviruses. *Annu. Rev. Microbiol.* **22**, 359–390.
10. Hozák, P., Hassan, A. B., Jackson, D. A. & Cook, P. R. (1993). Visualization of replication factories attached to a nucleoskeleton. *Cell*, **73**, 361–373.
11. Jackson, D. A., Hassan, A. B., Errington, R. J. & Cook, P. R. (1993). Visualization of focal sites of transcription within human nuclei. *EMBO J.* **12**, 1059–1065.
12. Jackson, D. A., Balajee, A. S., Mullenders, L. & Cook, P. R. (1994). Sites in human nuclei where DNA damaged by ultra-violet light is repaired: visualization and localization relative to the nucleoskeleton. *J. Cell Sci.* **107**, 1745–1752.
13. Cook, P. R. (1999). The organization of replication and transcription. *Science*, **284**, 1790–1795.

14. Jackson, D. A., McCready, S. J. & Cook, P. R. (1981). RNA is synthesised at the nuclear cage. *Nature*, **292**, 552–555.
15. Herbert, K. M., Greenleaf, W. J. & Block, S. M. (2008). Single-molecule studies of RNA polymerase: motoring along. *Annu. Rev. Biochem.* **77**, 149–176.
16. Rippe, K. (2001). Making contacts on a nucleic acid polymer. *Trends Biochem. Sci.* **26**, 733–740.
17. Marenduzzo, D., Finan, K. & Cook, P. R. (2006). The depletion attraction: an underappreciated force driving cellular organization. *J. Cell Biol.* **175**, 681–686.
18. Marenduzzo, D., Micheletti, C. & Cook, P. R. (2006). Entropy-driven genome organization. *Biophys. J.* **90**, 3712–3721.
19. Jhunjhunwala, S., van Zelm, M. C., Peak, M. M., Cutchin, S., Riblet, R., van Dongen, J. J. *et al.* (2008). The 3D structure of the immunoglobulin heavy-chain locus: implications for long-range genomic interactions. *Cell*, **133**, 265–279.
20. Mateos-Langerak, J., Bohn, M., de Leeuw, W., Giromus, O., Manders, E. M., Verschure, P. J. *et al.* (2009). Spatially confined folding of chromatin in the interphase nucleus. *Proc. Natl Acad. Sci. USA*, **106**, 3812–3817.
21. Dinant, C., Luijsterburg, M. S., Hofer, T., von Bornstaedt, G., Vermeulen, W., Houtsmuller, A. B. & van Driel, R. (2009). Assembly of multiprotein complexes that control genome function. *J. Cell Biol.* **185**, 21–26.
22. Hopfield, J. J. (1974). Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc. Natl Acad. Sci. USA*, **71**, 4135–4139.
23. Kimura, H., Sugaya, K. & Cook, P. R. (2002). The transcription cycle of RNA polymerase II in living cells. *J. Cell Biol.* **159**, 777–782.
24. Darzacq, X., Shav-Tal, Y., de Turris, V., Brody, Y., Shenoy, S. M., Phair, R. D. & Singer, R. H. (2007). In vivo dynamics of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **14**, 796–806.
25. Pettijohn, D. E. (1996). The nucleoid. In (Neidhardt, F. C., Curtiss, R., Ingraham, J. L., Lin, E. C. C., Brooks Low, K., Magasanik, B., Reznifoff, W. S., Riley, M., Schaechter, M. & Umbarger, H. E., eds), pp. 158–166, ASM Press, Washington, DC.
26. Deng, S., Stein, R. A. & Higgins, N. P. (2005). Organization of supercoil domains and their reorganization by transcription. *Mol. Microbiol.* **57**, 1511–1521.
27. Bon, M., McGowan, S. J. & Cook, P. R. (2006). Many expressed genes in bacteria and yeast are transcribed only once per cell cycle. *FASEB J.* **20**, 1721–1723.
28. French, S. L. & Miller, O. L. (1989). Transcription mapping of the *Escherichia coli* chromosome by electron microscopy. *J. Bacteriol.* **171**, 4207–4216.
29. Jin, D. J. & Cabrera, J. E. (2006). Coupling the distribution of RNA polymerase to global gene regulation and the dynamic structure of the bacterial nucleoid in *Escherichia coli*. *J. Struct. Biol*, **156**, 284–291.
30. Wade, J. T., Struhl, K., Busby, S. J & Grainger, D. C. (2007). Genomic analysis of protein-DNA interactions in bacteria: insights into transcription and chromosome organization. *Mol. Microbiol.* **65**, 21–26.
31. Saier, M. H. (2008). The bacterial chromosome. *Crit. Rev. Biochem. Mol. Biol.* **43**, 89–134.
32. Jackson, D. A., McCready, S. J. & Cook, P. R. (1984). Replication and transcription depend on attachment of DNA to the nuclear cage. *J. Cell Sci. Suppl.* **1**, 59–79.
33. Sinden, R. R., Carlson, J. O. & Pettijohn, D. E. (1980). Torsional tension in the DNA double helix measured

with trimethylpsoralen in living *E. coli* cells, analogous measurements in insect and human cells. *Cell*, **21**, 773–783.

34. Matsumoto, K. & Hirose, S. (2004). Visualization of unconstrained negative supercoils of DNA on polytene chromosomes of *Drosophila*. *J. Cell Sci.* **117**, 3797–3805.

35. Marenduzzo, D., Faro-Trindade, I. & Cook, P. R. (2007). What are the molecular ties that maintain genomic loops? *Trends Genet.* **23**, 126–133.

36. Palstra, R. J., de Laat, W. & Grosveld, F. (2008). Beta-globin regulation and long-range interactions. *Adv. Genet.* **61**, 107–142.

37. Sexton, T., Umlauf, D., Kurukuti, S. & Fraser, P. (2007). The role of transcription factories in large-scale structure and dynamics of interphase chromatin. *Semin. Cell Dev. Biol.* **18**, 691–697.

38. Simonis, M. & de Laat, W. (2008). FISH-eyed and genome-wide views on the spatial organisation of gene expression. *Biochim. Biophys. Acta*, **1783**, 2052–2060.

39. Mitchell, J. & Fraser, P. (2008). Transcription factories are nuclear subcompartments that remain in the absence of transcription. *Genes Dev.* **22**, 20–25.

40. Palstra, R. J., Simonis, M., Klous, P., Brasset, E., Eijkelkamp, B. & de Laat, W. (2008). Maintenance of long-range DNA interactions after inhibition of ongoing RNA polymerase II transcription. *PLoS ONE*, **e1661**, 3.

41. Cléard, F., Moshkin, Y., Karch, F. & Maeda, R. (2006). Probing long-distance regulatory interactions in the *Drosophila melanogaster* bithorax complex using Dam identification. *Nat. Genet*, **38**, 931–935.

42. Kadauke, S. & Blobel, G. A. (2009). Chromatin loops in gene regulation. *Biochim. Biophys. Acta*, **1789**, 17–25.

43. Pombo, A., Jackson, D. A., Hollinshead, M., Wang, Z., Roeder, R. G. & Cook, P. R. (1999). Regional specialization in human nuclei: visualization of discrete sites of transcription by RNA polymerase III. *EMBO J.* **18**, 2241–2253.

44. Jackson, D. A., Iborra, F. J., Manders, E. M. M. & Cook, P. R. (1998). Numbers and organization of RNA polymerases, nascent transcripts and transcription units in HeLa nuclei. *Mol. Biol. Cell*, **9**, 1523–1536.

45. Faro-Trindade, I. & Cook, P. R. (2006). A conserved organization of transcription during embryonic stem cell differentiation and in cells with high C value. *Mol. Biol. Cell*, **17**, 2910–2920.

46. Eskiw, C. H., Rapp, A., Carter, D. R. F. & Cook, P. R. (2008). RNA polymerase II activity is located on the surface of ~87 nm protein-rich transcription factories. *J. Cell Sci.* **121**, 1999–2007.

47. Hozák, P., Sasseville, A. M.-J., Raymond, Y. & Cook, P. R. (1995). Lamin proteins form an internal nucleoskeleton as well as a peripheral lamina in human cells. *J. Cell Sci.* **108**, 635–644.

48. Xu, M. & Cook, P. R. (2008). Similar active genes cluster in specialized transcription factories. *J. Cell Biol.* **181**, 615–623.

49. Kosak, S. T., Scalzo, D., Alworth, S. V., Li, F., Palmer, S., Enver, T. *et al.* (2007). Coordinate gene regulation during hematopoiesis is related to genomic organization. *PLoS Biol.* **e309**, 5.

50. Bon, M., Marenduzzo, D. & Cook, P. R. (2006). Modeling a self-avoiding chromatin loop: relation to the packing problem, action-at-a-distance, and nuclear context. *Structure*, **14**, 197–204.

51. Cook, P. R. (2003). Nongenic transcription, gene regulation and action at a distance. *J. Cell Sci.* **116**, 4483–4491.

52. Müeller-Storm, H., Sogo, J. & Schaffner, W. (1989). An enhancer stimulates transcription in trans when attached to the promoter via a protein bridge. *Cell*, **58**, 767–777.

53. Jackson, D. A. & Cook, P. R. (1993). Transcriptionally-active minichromosomes are attached transiently in nuclei through transcription units. *J. Cell Sci.* **105**, 1143–1150.

54. Margaritis, T. & Holstege, F. C. (2008). Poised RNA polymerase II gives pause for thought. *Cell*, **133**, 581–584.

55. Hozák, P., Cook, P., Schöfer, C., Mosgöller, W. & Wachtler, F. (1994). Site of transcription of ribosomal RNA and intranucleolar structure in HeLa cells. *J. Cell Sci.* **107**, 639–648.

56. Morgan, G. T. (2002). Lampbrush chromosomes and associated bodies: new insights into principles of nuclear structure and function. *Chromosome Res.* **10**, 177–200.

57. Gall, J. G. & Murphy, C. (1998). Assembly of lampbrush chromosomes from sperm chromatin. *Mol. Biol. Cell*, **9**, 733–747.

58. Cook, P. R. & Marenduzzo, D. (2009). Entropic organization of interphase chromosomes. *J. Cell Biol.* **186**, 825–834.

59. Xu, M. & Cook, P. R. (2008). The role of specialized transcription factories in chromosome pairing. *Biochem. Biophys. Acta*, **1783**, 2155–2160.

60. Cook, P. R. (2001). pp. 352, J. Wiley and Sons, New York.

61. Roussel, P., André, C., Comai, L. & Hernandez-Verdun, D. (1996). The rDNA transcription machinery is assembled during mitosis in active NORs and absent in inactive NORs. *J. Cell Biol.* **133**, 235–246.

62. Miller, O. J. & Bakken, A. (1972). Morphological studies of transcription. *Acta Endocrinol. Suppl. (Copenhagen)*, **168**, 155–177.