Predicting three-dimensional genome structure from transcriptional activity

Peter R. Cook

We would like to be able to predict how genomes are folded in the cell from the primary DNA sequence. A model for the three-dimensional structure of all genomes is presented; it is based on the structure of the bacterial nucleoid, where RNA polymerases cluster and loop the DNA. Loops appear and disappear as polymerases initiate and terminate, but the microscopic structure is 'self-organizing' and, to some extent, predictable. At the macroscopic level, transcriptional activity drives pairing between homologous sequences, inactivity allows genome compaction, and the segregation machinery orients whole chromosomes.

© 2002 Nature Publishing Group http://www.nature.com/naturegenetics

We now know the genomic DNA sequences of various organisms, including humans, but little about how those genomes are folded in a cell. These foldings pack genomes tightly; human DNA has a contour length of roughly 2 m and must fit into a nucleus with a diameter of roughly 10 μm. Several interrelated properties of genomes (that is, the DNA and associated proteins) further complicate analysis of their three-dimensional (3D) structure. First, genomes are so long and fragile that some breakage inevitably occurs during isolation, and the released DNA strands soon aggregate into an intractable gel. Second, aggregation is promoted by the high cellular concentration and charge of those genomes. Naturally, biochemists have developed purification procedures that minimize aggregation, but these usually employ hyper- or hypotonic buffers that distort structure[1]. Third, many folds have dimensions below the resolution of the light microscope (roughly 200 nm) and so can only be seen by electron microscopy, but this brings further problems associated with preserving structure *in vacuo*.

Given these difficulties, it is not surprising that details of 3D structure have been controversial. For example, the shape of the bacterial nucleoid seen in electron micrographs critically depends on fixation conditions[2]. And although all agree that in eukaryotes the double helix is coiled locally around nucleosomes and globally into discrete nuclear 'territories', the foldings in between are still under discussion[3]. Models for such foldings involve (i) random packing, in which nucleosomal strings are packed like spaghetti, perhaps generating loops of about 1 Mb[4]; (ii) helical hierarchies, in which strings are coiled into solenoids (of roughly 30 nm diameter), solenoids into higher-order structures and so on[5]; (iii) loops (50–150 kb) attached to the peripheral lamina or internal structures, such as (iv) 'skeletons'/'scaffolds'[6] or 'factories'[7]; and (v) combinations of the above—for example, of helical coils and radial loops[8] or helical coils and random folding[9]. Most models are specific to eukaryotes and cannot easily be extended to bacteria. Here, I describe a general model that applies to all genomes; it is based on observations made in the 1970s.

When bacterial cells are lysed in a detergent and 1 M NaCl, most proteins are stripped off the genome to leave clusters of still-engaged polymerases and their transcripts attached to and surrounded by 'halos' of naked superhelical DNA[2] (Fig. 1*a*). Since inhibition of transcription with rifampicin or treatment with ribonuclease (RNase) unfolds these nucleoids, active tran-

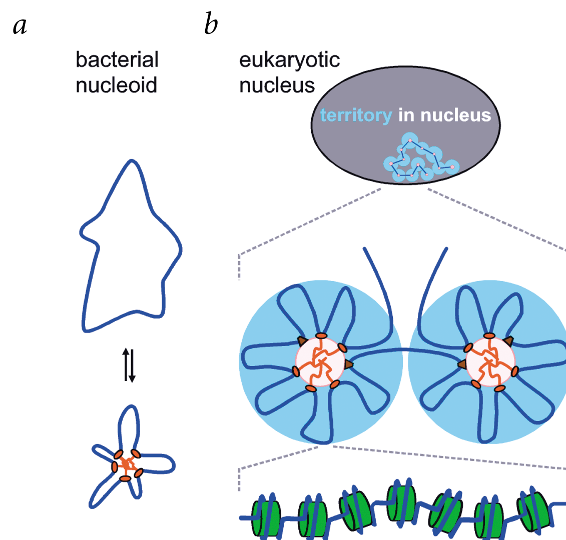


Fig. 1 Models of genome structure. ***a***, Prokaryotic nucleoid. Transcription of the circular chromosome (top), followed by aggregation of polymerases (ovals) and transcripts (red lines), generates a looped structure (bottom) that is self-sustaining (as promoters in active genes now lie close to polymerases). ***b***, Eukaryotic (HeLa-cell) nucleus. DNA is coiled around a histone octamer, and runs of nucleosomes form a zigzagging string (bottom). At the intermediate level in the structural hierarchy (middle), this string is organized into loops (average 86 kb; range 5–200 kb) by attachment to factories (pink circles) through transcription factors (trapezoids) and engaged RNA polymerases (ovals). Ten to twenty such loops (only a few are shown) form a chromatin cloud around the factory, which is equivalent to the structure of the bacterial nucleoid. Each cloud contains only one type of RNA polymerase (that is, I, II or III; refs 20,49), and some clouds are richer in housekeeping genes[50] or certain transcription factors[28] than others. Fifty to two hundred successive clouds (blue circles) strung along the chromosome form a territory (top; the general path of DNA between clouds is shown). At any moment, each chromosome possesses a unique array of transcription units and string of clouds; only homologs share roughly similar arrays.

Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK. Correspondence should be addressed to the author (e-mail: peter.cook@path.ox.ac.uk).

scription complexes maintain their structure. Similar isolates (that is, nucleoids) made from transcriptionally active eukaryotic cells (for example, fibroblasts, erythroblasts, epithelial cells of humans, chickens, frogs and insects) also contain superhelical loops surrounding a core of polymerases and their transcripts[10]. Notably, individual genes associate with the core only when they are transcribed; when inactive, they are out in loops. Moreover, looping is lost progressively as chicken erythroblasts mature; the inactive end cell—the erythrocyte—yields dispersed, unstructured DNA. All these results point to a structural role for the clustered and active polymerases, but this has never been widely accepted. The associations observed *in vitro* could have been generated artifactually, and the clustering implied that the active polymerases were immobilized, which clashed with a perceived need for them to track along the template as they made RNA. But recent evidence, derived from work on living cells and on isolates made in 'physiological' buffers that retain the polymerizing activity found *in vivo*, supports the idea that transcription maintains structure. In one sense, the idea that genome structure is a direct consequence of transcriptional activity is a tautology, as the shape of all cellular structures must depend on past and present transcription to generate the molecules of RNA and protein in those structures. But this model involves something much more immediate: the physical structure of the transcription complex is the critical organizer. I begin by discussing some design principles that constrain all models for genome structure.

Some design principles

Structural information in DNA is universally recognizable. Several models for genome structure involve looping mediated by proteins binding to DNA repeats. We might expect such protein and DNA motifs to be highly conserved, but data from the various genome projects has not shown this. Moreover, segments of bacterial and yeast DNA can integrate into mammalian DNA and be folded correctly into interphase and mitotic structures[11]. This implies that the integrated DNA contains the required structural information, and that mammalian proteins can interpret this information correctly. In other words, the critical interactions cannot be kingdom-specific. I will argue that such interactions stem from the ability of RNA polymerase to initiate somewhere in a DNA sequence to give a transient but stable elongation complex.

Self-organization. Macromolecular structures are generated in two fundamentally different ways[12] (Fig. 2*a*). Many virus particles 'self-assemble' according to a fixed plan to attain a true thermodynamic equilibrium; the particles are stable and static, and can survive in the absence of a pool of unincorporated subunits once they have been released from the host. But most cellular structures (for example, the cytoskeleton) are built using different principles. They lack a rigid architecture and are self-organizing. They are also intrinsically unstable, persisting only by exchanging subunits with others in their surroundings; if those subunits are removed, they collapse and eventually disappear. Structure depends on continuing function and vice versa. For example, when pure tubulin, microtubule motors and ATP are combined *in vitro*, the relative concentrations of the different components determine whether random networks, vortices or asters form; moreover, slight alterations in concentration can switch one pattern to another[13]. I will argue that transcription drives the self-organization of the genome. As self-organizing structures are dynamic, with their shape at a particular moment depending on past and present environments, statements about position cannot be precise; we should, however, be able to predict that a gene is more likely than not to be in a particular place, if we know the rules. (The term 'self-organizing' has different meanings in different scientific disciplines; it is used here in the dynamic sense just described.)

Molecular crowding promotes aggregation. Biochemists prefer to study reactions in dilute solutions where specific interactions dominate. In living cells, however, those reactions take place in the presence of substantial concentrations (0.1–0.4 g ml^{-1}) of macromolecules, some of which are parts of enormous insoluble arrays (for example, DNA or components of the cytoskeleton). As a result, much of the surrounding volume is occupied by macromolecules (Fig. 2*b*), and non-specific interactions involving them contribute considerably to total free energy. Although some biochemists regard such non-specific interactions as artifacts because they prevent the acquisition of meaningful data, nature inevitably exploits them.

Molecular crowding affects concentrations, equilibrium constants and reaction rates[14]. Consider the volume that one macromolecule prevents others from occupying (Fig. 2*c*). The effective concentration (or thermodynamic activity) of a large

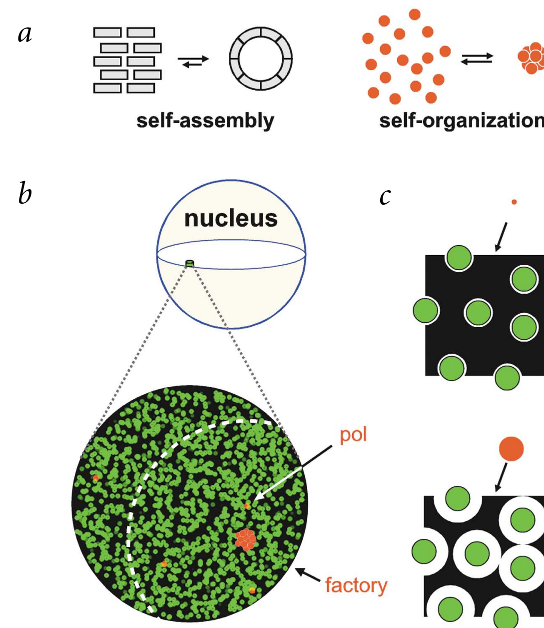


Fig. 2 Some design principles. ***a***, In self-assembly, components assemble into a stable, static structure that reaches thermodynamic equilibrium. In self-organization, components in the complex exchange continuously with those in a soluble pool; structure depends on function and vice versa. ***b***, Molecular crowding in the nucleus. A simulation of the nucleosomes (green circles), RNA polymerase II molecules (pol) and transcription factories in a cylindrical region (diameter 500 nm, height 50 nm) of a tetraploid HeLa nucleus. Molecules are drawn to scale at appropriate concentrations (except for the factory, which would be seen in only one of four such sections). Assumptions are that 6×10^7 nucleosomes and 3.2×10^5 polymerases (half spread diffusely, half in 8,000 factories) are distributed randomly in the nucleoplasm (volume 400 μm^3), and a typical chromatin loop (roughly 86 kb) extends approximately 250 nm (white line) from the factory. Non-histone proteins, RNA, ribonucleoproteins and internucleosomal DNA are each present at roughly the same concentration as nucleosomes, but are not shown; therefore, crowding is roughly five times that shown here. Even so, a nucleosome in a loop can diffuse freely within the space confined by the white line (covering 100 nm in roughly 15 s) before being constrained by those in neighboring chromatin clouds[51]. Inhibiting transcription also increases the mobility of loops[19]; presumably, transcription units disengage to create larger, more mobile loops. ***c***, Molecular crowding and effective concentration. The center of mass of the small red sphere (top) can only access the black volume, and not the volume occupied by the green spheres or the surrounding white volume that extends outwards by the radius of the red sphere. In contrast, the large red sphere (bottom) can access less of the same volume, simply because the white exclusion zone extends farther outwards. Therefore, the effective concentration, or thermodynamic activity, of the large red molecule is higher than that of the small one, even though the actual concentrations of the two are the same.

© 2002 Nature Publishing Group http://www.nature.com/naturegenetics

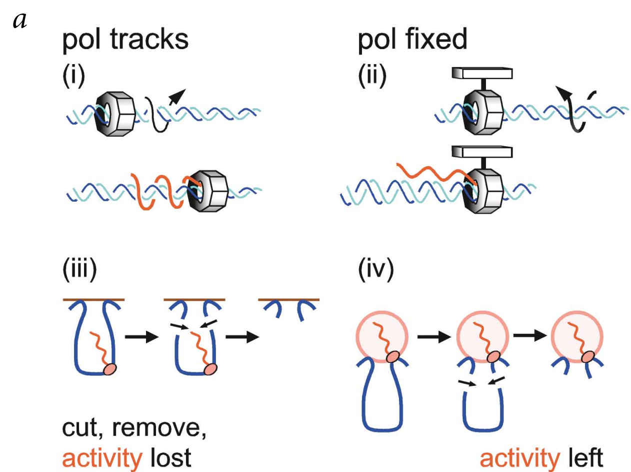
molecule is higher than that of a small one, even when the actual concentrations of the two are the same. For this reason, the effective concentration of hemoglobin in a red blood cell (measured from the osmotic pressure exerted) is about 100 times greater than the actual concentration of roughly 0.3 g ml^{-1}. The difference between effective and actual concentrations feeds through to influence equilibria: the effective equilibrium constant governing the reversible dimerization of a spherical protein of 100 kD is roughly 100 times higher in a red cell than at infinite dilution. This means that if a molecule tends to aggregate in a dilute solution, it is even more likely to do so in a crowded cell[14]. Biochemists find that complexes between DNA and the polymerase and between DNA and the most abundant DNA-binding proteins in bacteria and eukaryotes—HU proteins and histones, respectively—are all prone to aggregate in dilute isotonic solutions; therefore, they generally study them at higher or lower ionic strengths. For example, pure bacterial RNA polymerase sediments as an aggregate at 20–25S in an isotonic buffer, but as a monomer at 12.5–13.5S in ionic strengths above 0.25 (ref. 15), and where crowding becomes acute (for example, in viral particles or starved bacteria), DNA precipitates into a semi-crystalline array[16]. I will argue that the aggregation of transcription complexes strung along the genome inevitably folds that genome (Fig. 1*a*), and where there is no transcription, the effects of molecular crowding will collapse the genome into the smallest possible volume.

Genome structure depends on continuing transcription. In living bacteria, the nucleoid is found in the interior of the cell, but is dispersed when transcription is inhibited with rifampicin[17]. In eukaryotic cells, DNA sequences originally confined to distinct territories (or nucleoli or chromocenters) also disperse when mRNA synthesis is inhibited by 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB) or α-amanitin[18]; DRB also increases the mobility of loops in living cells[19].

DNA loops formed by clustered polymerases

Consideration of these principles leads to a general model[7] (Fig. 1). Polymerases bind to (and transcribe) transcription units scattered along the genome, and then active transcription complexes aggregate, forming the DNA into a surrounding halo or 'cloud' of loops. In bacteria, there are few such clouds; in eukaryotes, many. Once the stable attachments between an engaged polymerase, its transcript and template are lost on transcriptional termination, a new, larger loop is generated and the polymerase becomes free to exchange with others in solution. Notably, the released transcription unit is still near a polymerase cluster, favoring re-engagement. Thus, whereas the microscopic structure changes from moment to moment, the macroscopic structure is self-organizing and, to some extent, predictable. Active transcription units are likely to be associated with a cluster, units located adjacent to each other in the primary DNA sequence tend to attach to the same cluster, groups of active units separated by a long stretches of inactive DNA aggregate into separate clusters, and inactive DNA in between will be out in a loop. Then, each bacterial cell in a culture (or each eukaryotic cell at the same stage in a developmental pathway) would contain roughly the same linear array of active transcription units strung along the chromosome, and roughly the same number of polymerase clusters (or 'factories'). But a specified gene might be out in a loop at one moment and attached the next, and the precise attachments around that gene in another cell would rarely be the same. When the bacterial cells adapt as the culture conditions change (or when eukaryotic cells differentiate), a different constellation of loops forms. An exquisite functional order underlies the apparent chaos, however; in each population, a cluster with the appropriate polymerizing machinery is usually within reach of a potentially active gene.

In this model, active polymerases are immobilized and clustered. This notion clashes with the widely held assumption that polymerases track along their helical templates as they make RNA. But is this assumption correct? There is growing evidence that it is not[20]. There seems to be no direct evidence to support it, and immobilized polymerases work as efficiently as their soluble counterparts *in vitro*[21]. Logic also suggests that only immobilized polymerases can make transcripts that are free to escape to the cytoplasm (Fig. 3*a*), and models involving immobilized polymerases are now coming into favor[20,22]. Moreover, engaged enzymes resist nucleolytic removal from nuclei, suggesting that



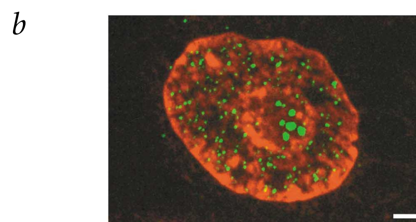


Fig. 3 Active polymerases are immobilized and clustered. ***a***, The untwining problem (i) and a solution (ii), and evidence that active polymerases are attached (iii, iv). (i) A tracking polymerase (shown here as a nut traveling along a double-helical bolt; the black arrow illustrates the track) generates a transcript (red) that is entwined about the template, but no satisfactory mechanism for untwining the transcript has yet been suggested. (ii) When a fixed polymerase reels in its template and extrudes its transcript, the untwining problem does not arise (the black arrow illustrates the motion of the template). In both cases, torsional stress accumulates (not shown) and is removed by topoisomerases[52]. Other formal possibilities are discussed elsewhere[20]. (iii) A polymerase (oval) tracks along a chromatin loop as it generates a transcript (left). When HeLa cells are permeabilized in a physiological buffer and treated with a nuclease (arrows; middle) and detached chromatin fragments are removed, the tracking polymerase should be lost with the detached chromatin (right). (iv) If the polymerase is attached to a factory, the engaged polymerase and its transcript should be retained. This is the result observed, and the kinetics of chromatin detachment are consistent with the existence of loops of 7.5–175 kb (averaging roughly 86 kb)[1]. ***b***, Nascent transcripts are concentrated in discrete foci in a human nucleus. HeLa cells were permeabilized, nascent transcripts were extended in 5-bromo-UTP, and cryosections (100 nm) were prepared. The Br-RNA was immunolabeled with FITC (green), nucleic acids were counterstained with TOTO-3 (red), and a fluorescence image was collected with a confocal microscope[49]. Heterochromatin (marked by intense red fluorescence) is concentrated around the nuclear periphery and the nucleolus in the interior, whereas nascent transcripts (green) are found in discrete foci in the cytoplasm (where they are made by the mitochondrial polymerase), nucleolus (the most intense foci) and nucleoplasm. Under these conditions, essentially all transcription sites are detected, and, as there are more active polymerases than foci, each focus must contain a cluster of active polymerases and their transcripts. Image provided by A. Pombo. Bar: 2 μm.

© 2002 Nature Publishing Group http://www.nature.com/naturegenetics

© 2002 Nature Publishing Group http://www.nature.com/naturegenetics

they must be attached (Fig. 3*a*). Decisively, polymerases tagged with the green fluorescent protein (GFP) are clustered in living *Bacillus subtilis* in discrete foci, or factories[23]. Each factory contains hundreds of RNA polymerase molecules engaged on some or all of seven *rrn* operons (calculated using data from *Escherichia coli*[24]). These operons, which encode rRNAs, are usually the most active in the cell; seven are spread over roughly 200 kb (contour length approximately 70 µm), but any associated polymerases are packed into less than 400 nm. Starvation reduces the transcription rate and disperses the polymerases. In eukaryotes, nascent transcripts are also clustered in factories containing many polymerases (Fig. 3*b*).

Studies of RNA polymerase I have given us a detailed view of a eukaryotic transcription factory dedicated to the production of 45S rRNA[20,25]. Some of these rRNA genes are the most active in the cell, and they organize the one subnuclear structure known to biologists of the pre-molecular era—the nucleolus. A triploid HeLa cell contains roughly 540 45S rRNA genes arranged in tandem repeats on different chromosomes; soon after mitosis, roughly 120 genes on some of these chromosomes become active, and only these nucleate approximately 30 'fibrillar centers'. Each of the 4 or so active genes associated with one center is transcribed by roughly 125 polymerases, and the resulting transcripts are found in one of the bright foci in the nucleolus (Fig. 3*b*). Polymerase II generates most messenger RNAs and is found in roughly 8,000 nucleoplasmic factories (Fig. 3*b*); these factories (approximately 50 nm in diameter) typically contain about eight active polymerases, each engaged on a different transcription unit[20,26].

Clustering ensures that the local concentration of polymerases and promoters in and around a factory is high (Fig. 2*b*). If the soluble molecules of polymerase II in a HeLa cell were spread randomly throughout the nucleus, they would be present at a concentration of roughly 1 µM (spaced approximately every 120 nm); inactive promoters are present at roughly the same concentration (and spacing). But the promoter of an active gene is inevitably tethered close to a factory, where the local concentration of polymerases is 1,000 times higher. Moreover, the promoter concentration falls off with the cube of the distance from the factory. Thus, active loops of roughly 10 kb[1] are restricted to the volume within 40 nm of the surface, but loops of average size (that is, 86 kb; ref. 1) can extend all the way to the next chromatin cloud (Fig. 2*b*). As a result, promoters in short loops are more likely to collide productively with the relevant polymerases and transcription factors in factories, especially when those in longer loops are immobilized in heterochromatin. These structures are inevitably self-sustaining; productive collisions tether a template that now attracts more of the relevant factors, increasing the chances that the next collision will also be productive.

Other principal factors affecting position

The segregation machinery in living bacteria orients the chromosome by pulling on daughter origins and leaving termini in the center[27] (Fig. 4). In eukaryotes, the spindle also pulls centromeres to the poles during anaphase, and when the nuclear membrane re-forms, it often does so around chromosomes that retain this polarized orientation (Fig. 4).

Transcription of repeated rRNA genes provides a second powerful influence on chromosome position. In living *B. subtilis*, it organizes the *rrn* operons into foci, and in eukaryotes, it brings together whole chromosomal territories[25] (Fig. 4). Genes for 45S rRNA are encoded by nucleolar organizing regions (NORs) carried on several human chromosomes. When chromosomes decondense after mitosis, one or more NORs may nucleate the formation of a nucleolar transcription factory. As a result of random chromosomal movements, a 45S rRNA promoter on one chromosome may collide with and initiate in a factory organized by another, so that those two chromosomes are now tethered to the same nucleolus. Here again, function generates a structure that further enhances function. Oct1/PTF/transcription domains also bring particular genes (and flanking chromosomal territories) to a region where the appropriate transcription and processing factors are concentrated[28], and coiled bodies may act in a similar manner[29].

Transcription may also drive pairing between homologous chromosomes[30] (Fig. 4). Though the condensed structures that we call chromosomes are usually transcriptionally inactive, when those structures pair they are always transcriptionally active. For example, condensed homologs remain active as they pair during meiotic prophase I. Notably, where a meiotic 'pairing' sequence has been mapped, it turns out to be a promoter, with promoter copy number determining the degree of pairing[31]. Somatic homologs also pair, and the degree of pairing correlates inversely with genome complexity—yeast homologs do so often, human ones rarely. This is probably because the pairing rate depends on the time it takes to make and break inappropriate contacts before a chromosome chances upon its homolog. At an intermediate level of complexity in the *Drosophila melanogaster* embryo, homologs pair only when transcription begins at the midblastula transition, and the active and repeated histone genes are some of the first to pair[32]; moreover, the giant polytene chromosomes seen later in development are both active and paired.

Transcription could drive pairing as follows[30]. Each chromatin fiber has a unique array of transcription units (and clouds) strung along it (Fig. 1*b*); some clouds are rich in long heterochromatic loops, others in short active loops tied to factories rich in particular factors. Only homologs share similar strings, although those strings differ in microscopic detail. Pairing could be nucleated when a promoter in one string initiates in a homologous factory containing the appropriate factors on the other. The two factories then remain tied together until termination, and this gives time for neighboring promoters to attach to homologous factories on the 'wrong' string; eventually, the two strings become zipped together.

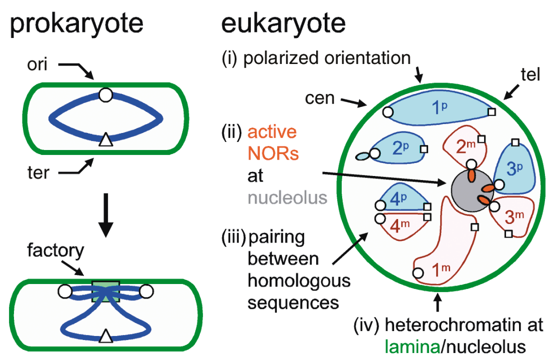


Fig. 4 Principal factors affecting position. In *E. coli*, the origin of replication (ori) is found centrally (top); after duplication in a replication factory, the segregation machinery pulls daughter origins to poles leaving the unreplicated terminus (ter) in the middle. Various factors position chromosome territories in a diploid eukaryotic nucleus (only four pairs of chromosomes are shown). (i) During anaphase, the spindle pulls centromeres (cen) to poles, leaving telomeres (tel) trailing behind; when the nuclear membrane reforms, it has done so around chromosomes 1^p, 2^p, 4^p, and 4^m that retain this polarized orientation. (ii) Four chromosomes bearing terminal NORs (ovals) are shown, but only three NORs (red ovals on 2^m, 3^p, 3^m) organize a nucleolus where they are transcribed; the other (blue oval on 2^p) remains unassociated and untranscribed. (iii) Some sequences on maternal and paternal homologs (that is, 4^p, 4^m) pair. (iv) Heterochromatin self-associates, often at the lamina and nucleolus (as rDNA genes are embedded in heterochromatin).

Inactivity also influences chromosome position; it promotes nucleosomal aggregation into heterochromatin, which in turn tends to compact on to the nuclear lamina and nucleoli[33] (Fig. 4). Here, the lack of function transforms the structure into one that further limits function.

Final gene position will then depend on a resolution of these principal influences, as well as many minor ones that have not been discussed. As a result, origins and termini of replication in bacteria—and centromeres, telomeres, NORs and heterochromatin in eukaryotes—will be more likely to be found in certain places than others, and they will tend to carry adjacent genes to those places.

Some problems and questions

Electron microscopy reveals RNA polymerases at the periphery of the bacterial cell, not in the center. Fixation must generate this distribution[2], as polymerases tagged with GFP are found centrally in living bacteria[23].

Active genes tend to lie on the surface of chromosome territories, not centrally, in eukaryotes. Only a few genes in mammalian cells showed this slight bias[34], and we now know that nascent transcripts (detected after extension in 5-bromo-UTP and immunolabeling) are spread throughout territories[35].

What maintains the organization of heterochromatin, which is transcriptionally inert? Heterochromatin has a range of transcriptional activities. In chicken erythrocytes, it is completely inactive, and electron microscopy reveals that its nucleosomes are tightly packed. But when erythrocyte genomes are isolated in 2 M NaCl, they lack any structure. Heterochromatic regions that are more structured turn out not to be as inactive as previously thought. For example, many genes in the facultative heterochromatin of the 'inactive' X chromosome of female mammals are active[36], as are many sequences in the constitutive centromeric heterochromatin of wheat[35]. Therefore, differences between euchromatin and heterochromatin probably stem from differences in loop length; the further a gene is from a factory, the less likely it is to be transcribed and the more likely it is to aggregate into heterochromatin. There are many other differences between the two types of chromatin (for example, degree of acetylation, phosphorylation and methylation[37] and the presence of proteins such as HP1; ref. 33), and each probably complements the others to pack the nucleosomes ever more tightly, thus making it less likely that they can access a factory.

Are associations between polymerases and transcription units the only ones responsible for maintaining structure? Studies of nuclear substructures done using non-physiological buffers have yielded many insights into nuclear structure; unfortunately, they have also generated much of the controversy concerning the molecular interactions involved[38]. For example, repeated motifs such as scaffold and 'matrix' attachment regions bind eukaryotic DNA to isolated substructures, but such attachments are not observed under isotonic conditions[39]. Moreover, DNA topoisomerase II, at one time perhaps the strongest candidate for the critical organizing protein, has now been eliminated from contention; photobleaching of living cells shows that the entire complement of GFP–topoisomerase exchanges with the soluble pool too rapidly for it to have a structural role[40]. It has yet to be shown whether other candidates such as locus control regions (LCRs), 'boundary' elements and 'silencers' are attachment points. Note, however, that some LCRs (for example, that of β-*globin*[41]) and boundary elements (for example, *scs*' next to *hsp70* in *D. melanogaster*[42]) are transcribed, and the resulting attachments could alter the distance between neighboring promoters and factories. Many other factors (such as cohesins, topoisomerases, chromatin remodeling complexes, histone acetylases and deacetylases) not discussed here also affect chromatin organization[33,37].

Which molecular interactions maintain the polymerase cluster? Polymerase–polymerase, transcript–transcript and polymerase–transcript interactions may all be involved; the bacterial polymerase self-aggregates *in vitro*[15], and RNase disrupts the structure of both the isolated bacterial nucleoid[2] and nuclear matrix[43].

What maintains the distinct shape of (inactive) mitotic chromosomes? Discussion here has concentrated on the eukaryotic genome during interphase, but this model is easily extended to mitosis. In that case, the contour length of the chromatin loops, and the basic shape of the chromatin clouds remain unchanged[1,44] so that decreased transcription, plus increased cloud–cloud and factory–factory aggregation, could drive self-organization into the most compact and stable structure, a cylinder of nucleosomes around an axial factory-based core[7].

Predicting 3D structure from primary sequence

We can predict 3D structure if we know the location of all transcription units in the primary sequence and their relative activities. Unfortunately, knowledge of both is incomplete. For example, we do not yet know the location of all genes in the human genome, let alone all the transcription units. And although we have inventories of the different poly(A)$^+$ mRNAs in a cell (obtained using microarrays), these are related only indirectly to the corresponding inventories of primary transcripts. This follows logically because rates of production, processing, export and turnover of most mRNAs are unknown, and because these microarray approaches do not analyze nongenic poly(A)$^-$ transcripts. We also know little about the rules governing chromosome pairing and segregation or aggregation into heterochromatin. Predictions about any self-organizing structure are necessarily probabilistic rather than absolute. Nevertheless, the models in Fig. 1 are testable, and some possible approaches include: (i) transcription factories in living cells can be visualized in sharper detail in bacterial[23] and mammalian[45] cells expressing GFP-tagged polymerases using new high-resolution imaging techniques[46]; (ii) the distance in 3D nuclear space between genes strung along a chromosome can be mapped using the 'chromosome conformation capture' assay[47] (to see if distant active genes become locally concentrated in factories); and (iii) single-cell gene expression profiling allows simultaneous localization of approximately ten different nascent transcripts[48] (and whether they, too, are locally concentrated). Therefore, I hope the general principles outlined above will eventually enable us to deduce the chances that a gene will be found in one place rather than another.

Acknowledgments

I thank the Wellcome Trust for support and A. Pombo and M. Lloyd for providing figures and software.

Received 5 April; accepted 10 September 2002.

1. Jackson, D.A., Dickinson, P. & Cook, P.R. The size of chromatin loops in HeLa cells. *EMBO J.* **9**, 567–571 (1990).
2. Pettijohn, D.E. in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology* (eds Neidhardt, F.C., Curtiss, R., Ingraham, J.L., Lin, E.C.C., Low, K.B., Magasanik, B., Reznifoff, W.S., Riley, M., Schaechter, M. & Umbarger, H.E.) 158–166 (ASM Press, Washington, 1996).
3. Cook, P.R. *Principles of Nuclear Structure and Function* (Wiley, New York, 2001).
4. Sachs, R.K., van den Engh, G., Trask, B., Yokota, H. & Hearst, J.E. A random-walk/giant-loop model for interphase chromosomes. *Proc. Natl Acad. Sci. USA* **92**, 2710–2714 (1995).
5. Sedat, J. & Manuelidis, L. A direct approach to the structure of eukaryotic chromosomes. *Cold Spring Harb. Symp. Quant. Biol.* **42**, 331–350 (1978).
6. Saitoh, Y. & Laemmli, U.K. From the chromosomal loops and the scaffold to the classic bands of metaphase chromosomes. *Cold Spring Harb. Symp. Quant. Biol.* **58**, 755–765 (1978).
7. Cook, P.R. A chromomeric model for nuclear and chromosome structure. *J. Cell Sci.* **108**, 2927–2935 (1995).
8. Manuelidis, L. A view of interphase chromosomes. *Science* **250**, 1533–1540 (1990).
9. Li, G., Sudlow, G. & Belmont, A.S. Interphase cell cycle dynamics of a late-replicating, heterochromatic homogeneously staining region: precise

© 2002 Nature Publishing Group http://www.nature.com/naturegenetics

choreography of condensation/decondensation and nuclear positioning. *J. Cell Biol.* **140**, 975–989 (1998).
10. Jackson, D.A., McCready, S.J. & Cook, P.R. Replication and transcription depend on attachment of DNA to the nuclear cage. *J. Cell Sci. Suppl.* **1**, 59–79 (1984).
11. McManus, J. *et al*. Unusual chromosome structure of fission yeast DNA in mouse cells. *J. Cell Sci.* **107**, 469–486 (1994).
12. Misteli, T. The concept of self-organization in cellular architecture. *J. Cell Biol.* **155**, 181–185 (2001).
13. Nedelec, F., Surrey, T., Maggs, A.C. & Leibler, S. Self-organization of microtubules and motors. *Nature* **389**, 305–308 (1997).
14. Minton, A.P. The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. *J. Biol. Chem.* **256**, 10577–10580 (2001).
15. Ishihama, A. Subunit of assembly of *Escherichia coli* RNA polymerase. *Adv. Biophys.* **14**, 1–35 (1981).
16. Minsky, A., Shimoni, E. & Frenkiel-Krispin, D. Stress, order and survival. *Nature Rev. Mol. Cell Biol.* **3**, 50–60 (2002).
17. Mascarenhas, J., Weber, M.H. & Graumann, P.L. Specific polar localization of ribosomes in *Bacillus subtilis* depends on active transcription. *EMBO Rep.* **2**, 685–689 (2001).
18. Haaf, T. & Ward, D.C. Inhibition of RNA polymerase II transcription causes chromatin decondensation, loss of nucleolar structure, and dispersion of chromosomal domains. *Exp. Cell. Res.* **224**, 163–173 (1996).
19. Chubb, J.R., Boyle, S., Perry, P. & Bickmore, W.A. Chromatin motion is constrained by association with nuclear compartments in human cells. *Curr. Biol*. **12**, 439–445 (2002).
20. Cook, P.R. The organization of replication and transcription. *Science* **284**, 1790–1795 (1999).
21. Gelles, J. & Landick, R. RNA polymerase as a molecular motor. *Cell* **93**, 13–16 (1998).
22. Maniatis, T. & Reed, R. An extensive network of coupling among gene expression machines. *Nature* **416**, 499–506 (2002).
23. Lewis, P.J., Thaker, S.D. & Errington, J. Compartmentalization of transcription and translation in *Bacillus subtilis*. *EMBO J.* **19**, 710–718 (2000).
24. Bremer, H. & Dennis, P.P. in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology* (eds Neidhardt, F.C., Curtiss, R., Ingraham, J.L., Lin, E.C.C., Low, K.B., Magasanik, B., Reznifoff, W.S., Riley, M., Schaechter, M. & Umbarger, H.E.) 1553–1569 (ASM Press, Washington, 1996).
25. Shaw, P.J. & Jordan, E.G. The nucleolus. *Ann. Rev. Cell Dev. Biol.* **11**, 93–121 (1995).
26. Jackson, D.A., Iborra, F.J., Manders, E.M.M. & Cook, P.R. Numbers and organization of RNA polymerases, nascent transcripts and transcription units in HeLa nuclei. *Mol. Biol. Cell* **9**, 1523–1536 (1998).
27. Lemon, K.P. & Grossman, A.D. The extrusion-capture model for chromosome partitioning in bacteria. *Genes Dev.* **15**, 2031–2041 (2001).
28. Pombo, A. *et al.* Regional and temporal specialization in the nucleus: a transcriptionally-active nuclear domain rich in PTF, Oct1 and PIKA antigens associates with specific chromosomes early in the cell cycle. *EMBO J.* **17**, 1768–1778 (1998).
29. Frey, M.R., Bailey, A.D., Weiner, A.M. & Matera, A.G. Association of snRNA genes with coiled bodies is mediated by nascent snRNA transcripts. *Curr. Biol.* **9**, 126–135 (1999).
30. Cook, P.R. The transcriptional basis of chromosome pairing. *J. Cell Sci.* **110**, 1033–1040 (1997).
31. McKee, B.D., Habera, L. & Vrana, J.A. Evidence that intergenic spacer repeats of *Drosophila melanogaster* rRNA genes function as X–Y pairing sites in male meiosis, and a general model for achiasmatic pairing. *Genetics* **132**, 529–544 (1992).
32. Hiraoka, Y. *et al*. The onset of homologous chromosome pairing during *Drosophila melanogaster* embryogenesis. *J. Cell Biol.* **120**, 591–600 (1993).
33. Henikoff, S. Heterochromatin function in complex genomes. *Biochim. Biophys. Acta.* **1470**, 1–8 (2000).
34. Kurz, A. *et al*. Active and inactive genes localize preferentially in the periphery of chromosome territories. *J. Cell Biol.* **135**, 1195–1205 (1996).
35. Abranches, R., Beven, A.F., Aragon-Alcaide, L. & Shaw, P.J. Transcription sites are not correlated with chromosome territories in wheat nuclei. *J. Cell Biol.* **143**, 5–12 (1998).
36. Sudbrak, R. *et al*. X chromosome-specific cDNA arrays: identification of genes that escape from X-inactivation and other applications. *Hum. Mol. Genet.* **10**, 77–83 (2001).
37. Jenuwein, T. & Allis, C.D. Translating the histone code. *Science* **201**, 1074–1080 (2001).
38. Pederson, T. Half a century of 'the nuclear matrix'. *Mol. Biol. Cell* **11**, 799–805 (2000).
39. Jackson, D.A., Bartlett, J. & Cook, P.R. Sequences attaching loops of nuclear and mitochondrial DNA to underlying structures in human cells: the role of transcription units. *Nucleic Acids Res.* **24**, 1212–1219 (1996).
40. Christensen, M.O. *et al*. Dynamics of human DNA topoisomerases IIα and IIβ in living cells. *J. Cell Biol*. **157**, 31–44 (2002).
41. Gribnau, J., Diderich, K., Pruzina, S., Calzolari, R. & Fraser, P. Intergenic transcription and developmental remodeling of chromatin subdomains in the human β-*globin* locus. *Mol. Cell* **5**, 377–386 (2000).
42. Glover, D.M., Leibowitz, M.H., McLean, D.A. & Parry, H. Mutations in *aurora* prevent centrosome separation leading to the formation of monopolar spindles. *Cell* **81**, 95–105 (1995).
43. Nickerson, J.A. Experimental observations of a nuclear matrix. *J. Cell Sci.* **114**, 463–474 (2001).
44. Manders, E.M.M., Kimura, H. & Cook, P.R. Direct imaging of DNA in living cells reveals the dynamics of chromosome formation. *J. Cell Biol.* **144**, 813–822 (1999).
45. Sugaya, K., Vigneron, M. & Cook, P.R. Mammalian cell lines expressing functional RNA polymerase II tagged with the green fluorescent protein. *J. Cell Sci*. **113**, 2679–2683 (2000).
46. Gustafsson, M.G. Extended resolution fluorescence microscopy. *Curr. Opin. Struct. Biol*. **9**, 627–634 (1999).
47. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
48. Levsky, J.M., Shenoy, S.M., Pezo, R.C. & Singer, R.H. Single-cell gene expression profiling. *Science* **297**, 836–840 (2002).
49. Pombo, A. *et al.* Regional specialization in human nuclei: visualization of discrete sites of transcription by RNA polymerase III. *EMBO J.* **18**, 2241–2253 (1999).
50. Lercher, M.J., Urrutia, A.O. & Hurst, L.D. Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nat. Genet*. **31**, 180–183 (2002).
51. Marshall, W.F. *et al*. Interphase chromosomes undergo constrained diffusional motion in living cells. *Curr. Biol.* **7**, 930–939 (1997).
52. Liu, L.F. & Wang, J.C. Supercoiling of the DNA template during transcription. *Proc. Natl Acad. Sci. USA* **84**, 7024–7027 (1987).

© 2002 Nature Publishing Group http://www.nature.com/naturegenetics