Nongenic transcription, gene regulation and action at a distance

Peter R. Cook

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

Summary

In eukaryotes, motifs such as silencers, enhancers and locus control regions act over thousands of base pairs to regulate adjacent genes; insulators limit such effects, and barriers confine repressive heterochromatin to particular chromosomal segments. Recent results show that many of these motifs are nongenic transcription units, and two of them directly contact their targets lying further down the chromosome to loop the intervening DNA: the barriers (*scs* and *scs'*) flanking the 87A7 heat-shock locus in the fly contact each other, and a locus control region touches the β-globin gene in the mouse. I hypothesize that the act of transcription underlies the function of these regulators; active polymerizing complexes tend to cluster into 'factories' and this facilitates molecular contact between the transcribed regulator and its distant (and transcribed) target.

Key words: Barrier, Enhancer, Heterochromatin, Locus control region, Position effect, Silencer

Introduction

In bacteria, most regions regulating transcription lie within a few tens of base pairs of the transcriptional start site. They generally bind repressors and activators, and the bound regulators contact the transcription machinery to vary initiation rates up to 1000-fold. In higher eukaryotes, an additional but ill-defined nuclear 'context' acts over thousands of base pairs to regulate transcription by another 10,000-fold or more (Ivarie et al., 1983). Although a histone 'code' forms part of this context (Jenuwein and Allis, 2001; Kurdistani and Grunstein, 2003), there must be additional layers of control; yeast mutants lacking the N-terminal tails that carry the code are viable (Grunstein, 1990), and depleting histone H4 has little effect on the expression of most genes (Wyrick et al., 1999). Here, I discuss recent evidence showing that many regulatory motifs acting at a distance are transcription units, and that clusters of such units organize the chromatin fibre into loops. I suggest that the act of transcription of the regulatory motifs underlies the way they work and constitutes an under-recognized part of this context. I restrict discussion to a few of the best-characterized regulatory motifs, although many other examples could be cited.

An important distinction is made between genes (which usually encode proteins) and transcription units (which might be copied into noncoding transcripts). It is now estimated there are tenfold more transcription units than genes in the human genome (Kapranov et al., 2002), and very little is known about the former. Most of these extra transcription units do not encode recognizable structural RNAs like rRNA or tRNA, and most are not found in current transcriptomes [as they are copied into unstable transcripts that lack the poly(A) tails usually used to select RNA molecules for analysis]. A small fraction of these extra units will be pseudogenes, as sequence analysis reveals that many of the ~20,000 human pseudogenes have

functional promoters. [Note that transcription of one pseudogene (*Makorin1-1p*) regulates the expression of its coding counterpart (Hirotsune et al., 2003).] A larger fraction is probably made up of repeated sequences, because many of the retrotransposons in those repeats also contain functional promoters. However, most are probably just 'junk' and destroyed rapidly by the nonsense-mediated decay pathway (Iborra et al., 2001). I will argue that nature utilizes neighbouring transcription units (whether they be genic or nongenic) in gene regulation.

Heterochromatin: a repressive context

Heterochromatin was originally defined as the part of a mitotic chromosome like a centromere that remains condensed and dark-staining during interphase. It is characterized by the inaccessibility of its DNA to nucleases, methylation of cytosines in CpG doublets, methylation of lysine 9 in histone H3, a regular nucleosomal spacing, hypoacetylation of nucleosomes, and the presence of specific proteins, such as HP1 (Richards and Elgin, 2002). It contains few genes, and so was traditionally thought to be transcriptionally inert; however, we now know it contains many active transcription units. For example, wheat centromeres provide extreme examples of the deepest constitutive heterochromatin, and yet the density of transcription sites in this heterochromatin equals that of euchromatin (Abranches et al., 1998). [This was shown by permeabilizing cells in which centromeres and telomeres were found at opposite ends of nuclei, extending nascent transcripts in Br-UTP, and then immunolabelling the resulting Br-RNA.] In addition, many genes in the facultative heterochromatin of the so-called 'inactive' X chromosome of female mammals turn out to be active (Sudbrak et al., 2001), and the heterochromatin on the human Y chromosome contains many

transcription units (Skaletsky et al., 2003). Recent work even shows that transcription from within the heterochromatin of fission yeast is required to maintain the heterochromatic state (Volpe et al., 2002).

Heterochromatinization inactivates most genes (Gerasimova and Corces, 2001; Donze and Kamakaka, 2002; Schedl and Broach, 2003). During his famous experiments on the effects of X-rays on *Drosophila*, Muller noticed some mutants with mottled red and white eyes (Sturtevant and Beadle, 1962). This was subsequently traced to the translocation of the *white* gene from euchromatin to heterochromatin (Baker, 1968). Translocation – and not mutation of *white* – causes this variegation, because the phenotype reverts on relocation of the gene back to euchromatin. *white* encodes a membrane protein that transports pigment precursors into eye cells, and so deficient eyes appear shades of orange to white. *white* has a weak promoter without a canonical TATA box that is sensitive to adjacent sequences. Inactivation of *white* in some cells in the primordial eye disc early in development coupled to the subsequent inheritance of that inactivity leads to patches with and without pigment. Inactivation can be accompanied by the spread of heterochromatin hundreds of kilobases into juxtaposed euchromatin, with a consequential alteration in the banding pattern in the giant chromosomes of salivary glands. This spread propagates down one fibre and laterally to others through various modifications (e.g. histone hypoacetylation) until some 'barrier' is encountered (Donze and Kamakaka, 2002; Schedl and Broach, 2003). Genetic screens in *Drosophila* originally revealed that almost any euchromatic gene can be inactivated by such position effects (Lindsley et al., 1960; Baker, 1968), and this is now being amply confirmed in a wide range of higher eukaryotes by experiments using reporters encoding the green fluorescent protein (GFP) (Spector, 2003).

Current models for the way regulatory motifs act at a distance are of three non-exclusive types (Gerasimova and Corces, 2001; Schedl and Broach, 2003). They involve altering the balance between euchromatin and heterochromatin, and the motifs are seen as binding sites for proteins that induce (1) chemical modifications and/or structural alterations that propagate down the fibre, (2) looping to bring regulatory elements into contact with distant promoters, or (3) relocation of elements to permissive (or restrictive) nuclear compartments.

Transcription units as barriers, boundaries, insulators, enhancers, LCRs

I now describe examples of the best-characterized regulatory motifs; significantly, all prove to contain promoters and/or binding sites for transcription factors.

The first barrier segregating euchromatin from heterochromatin was defined in the *Drosophila* 87A7 heat-shock locus (Fig. 1A) (Kellum and Schedl, 1991). This locus possesses flanking 'specialized chromatin sequences' (*scs* and *scs'*) that each contain nuclease-hypersensitive sites and a nuclease-resistant core to which specific transcription factors bind (SBP/Zw5 binds to *scs*; BEAF32/DREF binds to *scs'*). When flies are transformed with wild-type *white*, some flies have wild-type eyes, and others have mutant ones. When *white* integrates into euchromatin it is expressed, and when integrated into heterochromatin it is repressed. But when it is flanked by
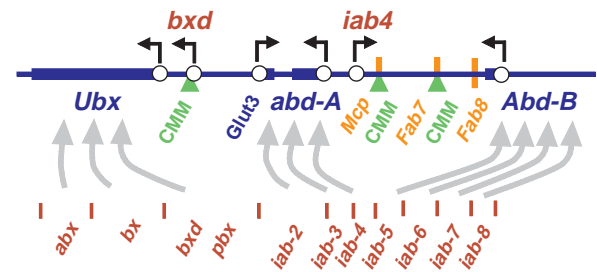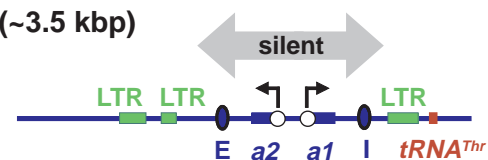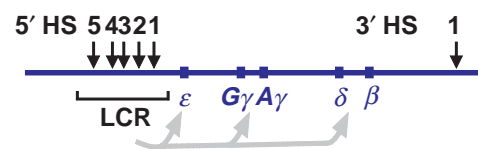

Fig. 1. Transcribed regulatory regions in four loci. (A) *Drosophila* 87A7 heat-shock locus. Two *hsp70* genes (promoters and direction of transcription indicated) are protected from the spread of heterochromatin (grey arrowheads) by *scs* and *scs'* (which both contain promoters). (B) *Drosophila* bithorax complex. *Ubx*, *abd-A* and *Abd-B* are regulated by many silencers and enhancers found in the regions shown below (grey arrows indicate target genes) and various cellular memory modules (CMMs); various barriers (e.g. *Mcp*, *Fab7*, *Fab8*) separate these regions. Not all ORFs, regulators, barriers or transcription units are shown. (C) Yeast HMR locus. A silent domain containing mating type (*MAT*) loci (*a2*, *a1*) is maintained by barriers flanking E and I. Rap1/Abf1/ORC bind to *ARS317* (at E) and *ARS318* (at I), followed by recruitment of SIR1-4, spread of SIR2-4 to flanking barriers and nucleosomal deacetylation by SIR2. The barrier next to I contains a long terminal repeat (LTR) and a *tRNA^{Thr}* gene; the one next to E probably contains one (or two) LTRs, and so also contains promoters. ORF YCR097W-A (downstream of *a1*) is not shown. (D) Human β-globin locus. A (transcribed) LCR regulates (grey arrows) the activity of the locus that contains five genes and one pseudogene (Ψβ; not shown). The vertical arrows mark hypersensitive sites (HS).

scs and *scs'*, most flies have wild-type eyes. Both *scs* and *scs'* contain promoters (Avramova and Tikhonov, 1999). Many similar barriers have now been identified (Gerasimova and Corces, 2001; West et al., 2002); hypersensitive site 4 (HS4) of chicken β-*globin* (see below) protects *white* from fly heterochromatin, and active transcription units insulate GFP-

encoding reporters from human heterochromatin (Sutter et al., 2003).

The *Drosophila* bithorax complex (BX-C) contains many well-studied enhancers, silencers and insulators that ensure that three homeotic genes – *Ubx*, *abd-A* and *Abd-B* – are expressed appropriately along the anterior-posterior axis of the developing fly (Fig. 1B) (Martin et al., 1995). A significant number of these regulators extend over many kilobases, and there has been no satisfactory explanation of why they are so long. Moreover, it has only recently been recognized that it is the act of transcription that underlies their activity. For example, first, deleting an unrelated gene (*Glut3*) encoding a glucose transporter that happens to be embedded in the locus alters *abd-A* expression and transforms the first abdominal segment into the second segment (Martin et al., 1995). Second, P element insertions downstream of *abd-A* also transform the first abdominal segment into the second and third segments; the P element promoter drives transcription across intergenic regions regulating segmental development, and blocking it reverts the mutant phenotype (Bender and Fitzgerald, 2002). Third, disrupting the transcription of each of the *iab-4* to *iab-8* enhancers transforms the segments they control into more posterior ones (Drewell et al., 2002a; Hogga and Karch, 2002). Fourth, activating transcription of the cellular memory modules (CMMs) associated with the *bxd*, *Mcp* and *Fab7* regulators correlates with the alteration in epigenetic state (Rank et al., 2002). Fifth, transcription of the *Mcp* and *Fab8* insulators correlates with their activity (Drewell et al., 2002a).

Saccharomyces cerevisiae lacks cytologically detectable heterochromatin (it is too small to be visible), but insertions in a few sites are subject to position effects. Their DNA is resistant to nucleases and their nucleosomes are hypoacetylated, but CpGs plus lysine 9 of histone H3 are unmethylated, and SIR (silent information regulator) proteins replace HP1. The most-studied region is *HMR*, which contains silent copies of the mating-type specific gene, *MATa* (Fig. 1C). A reporter gene such as *URA3* is silenced when inserted into *HMR* (Richards and Elgin, 2002). Silencing depends on context, because excised DNA rings soon reactivate (Cheng and Gartenberg, 2000). Barriers marking the limit of the silent domain contain promoters, and mutating the $tRNA^{Thr}$ promoter in the right-hand one reduces barrier activity (Donze and Kamakaka, 2001). [*tRNA* genes also limit the spreading of heterochromatin in fission yeast (Partridge et al., 2000).]

HS4, which lies 5′ of the chicken β-globin locus, is the best-characterized vertebrate boundary (Burgess-Beusse et al., 2002). It marks the border between nuclease-sensitive and nuclease-resistant chromatin, blocks enhancer action and screens reporters against position effects. Although it has not been shown to be a promoter in vivo, it is nevertheless a CpG island that bears the histone code (i.e. H3 hyperacetylation and methylation of lys4) characteristic of one (Litt et al., 2001).

Enhancers promote the activity of linked promoters (Hertel et al., 1997). They invariably contain binding sites for transcription factors and some are transcribed, including those regulating BX-C (above) and a subset known as locus control regions (LCRs). LCRs enable inserted transgenes to be expressed independently of position effects at physiological levels in a manner that is tissue specific and dependent on copy number (Li et al., 2002). The first LCR was found in the human β-globin locus (Fig. 1D). In transgenic mice that contain only the minimal β-*globin* gene, inserted transgenes were subject to position effects; however, adding distant hypersensitive sites (HS1-HS5) led to high-level expression in erythroid cells regardless of the insertion site. Both human and mouse LCRs contain multiple transcription units (Ashe et al., 1997; Routledge et al., 2002), and active RNA polymerase II can be immunoprecipitated bound to HS2 (Johnson et al., 2001). Other transcribed LCRs include those controlling the expression of α-globin, keratin 18, adenosine deaminase, growth hormone and major histocompatibility complex (MHC) class II genes (Li et al., 2002). Significantly, a point mutation in the promoter of the keratin 18 LCR – an *alu* repeat – destroys some of its insulating activity (Willoughby et al., 2000).

Transcription of all these motifs regulates the activity of linked genes. Other well-known mammalian examples include *H19* (a noncoding unit that combines with transcribed enhancers to control *IGF2* imprinting) (Drewell et al., 2002b), and *XIST* and *Tsix* (noncoding units maintaining inactivity of one of the X chromosomes in females) (Plath et al., 2002).

Repression through antisense transcription and transcriptional interference

One factor confounds the analysis of these motifs – the assays used alter flanking sequences, but the function examined depends on them. Thus, *scs* and *scs'* cannot insulate *white* from some heterochromatin on the X chromosome (Kellum and Schedl, 1991), *HMR* barriers protect one reporter (i.e. *URA3*) but not other genes with powerful upstream activation sequences (e.g. *TEF1* and *TEF2*) (Bi and Broach, 1999), *CHA1* – which flanks the *HML* mating-type locus – becomes a robust barrier when induced by serine (Donze and Kamakaka, 2001), and inverting the β-globin LCR destroys much of its activity (Tanimoto et al., 1999). I now argue that moving a motif from one place to another usually changes the flanking transcription units, and this generates much of the complexity seen.

Transcripts copied from noncoding strands can pair with their coding counterparts to regulate gene activity post-transcriptionally – for example, through the operation of the RNA interference, alternative splicing, and editing pathways (Gottesman, 2002; Storz, 2002). Such downstream effects are not discussed here. However, the act of antisense transcription inevitably and directly reduces sense transcription, because both strands cannot be copied simultaneously. Such antisense inhibition is probably significant as at least 300 human genes contain embedded antisense units that are copied into transcripts stable enough to yield expressed sequence tags (ESTs) (Shendure and Church, 2002).

Transcription of a gene also interferes with the expression of a nonoverlapping but neighbouring one. This kind of transcriptional interference was first seen in mammals in clones harbouring a single (integrated) copy of a retroviral vector encoding resistance to both neomycin and azaguanine (Emerman and Temin, 1986). Expression of the 3′ resistance gene was suppressed when selection required expression of the 5′ gene, and vice versa. Moreover, few cells grew in both neomycin and azaguanine, showing that only one gene is usually active. Interference can spread at least 10 kbp, but the upper limit remains unknown. It is generally missed in cells with many integrants (and in diploid cells) because one copy might express one gene at any moment, and a second copy

another. However, it is likely to be common in human cells, because ~20% of genes on chromosomes 21 and 22 lie within 1 kbp of another (in a head-to-head orientation) (Adachi and Lieber, 2002), and even more common when (unrecognized) transcription units are considered. Such interference also occurs between different kinds of unit; thus, *tRNA^Thr* (a polymerase III unit) silences *URA3* (a polymerase II unit; above), and other yeast *tRNA* genes reduce the transcription of polymerase II units lying within ~400 bp between three- and sixty-fold (Bolton and Boeke, 2003). Although we might imagine that two mammalian genes lying within a few kbp on a chromosome could each be loaded with many polymerases, this evidence suggests that only one of the two is transcribed at any moment.

Chromatin loops

An enduring idea in biology sees the chromatin fibre as looped, with attachment points as the barriers that segregate contexts. But despite wide acceptance, there remains little agreement as to which proteins and DNA motifs constitute the molecular ties. There is good reason for this. Chromatin is poised in a metastable state, and the buffers used for biochemical purification can irreversibly change the structure. Consequently, results obtained with isolates like 'matrices', 'scaffolds' and 'nucleoids' are not widely accepted (e.g. Belmont, 2002). Nevertheless, we can envisage two kinds of tie (Fig. 2) (Cook, 2001). 'Structural' ties would persist from one interphase to the next, and probably involve conserved DNA repeats. However, 30 years of research (including sequencing whole genomes) has failed to uncover any such
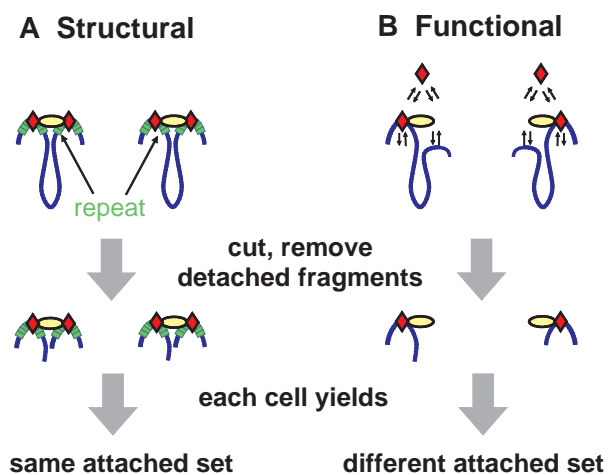
Fig. 2. Different loop ties. (A) Structural. DNA repeats (green) in two cells bind to the same protein complexes (yellow ovals, red diamonds), looping the fibre. After cutting with a nuclease and removing detached fragments, the same set of repeats from each cell remain bound. When 10% DNA remains, repeats are enriched tenfold. (B) Functional. The fibre is looped by attachment to a protein complex, but both attachments and proteins in the complex change from moment to moment. After cutting and removing detached fragments, a different set of fragments remains attached in the two cells. When 10% DNA remains, no fragment is enriched tenfold. This result is obtained if cutting and removal are carried out in isotonic buffers; essentially all active transcription complexes also remain attached.

motifs. 'Functional' ties would depend on which part of the genome was being transcribed, and even two daughter cells would possess different arrays of everchanging attachments. I believe we must now look to this elusive class.

Images of lampbrush chromosomes are often thought to provide the best evidence for looping. These chromosomes can be isolated from oocytes of many animals (but not mammals). During the first meiotic division, duplicated homologs pair, and long loops can be seen extending microns away from axial chromomeres. Unusually, these chromosomes are transcribed, and transcripts are attached to both loops and chromomeres (Snow and Callan, 1969). But these loops only become visible on dispersing chromatin in unphysiological buffers, and none are seen in sections of whole oocytes where chromatin appears as a granular aggregate; therefore, the long loops may be created during isolation as active units are stripped off chromomeres (Cook, 2001).
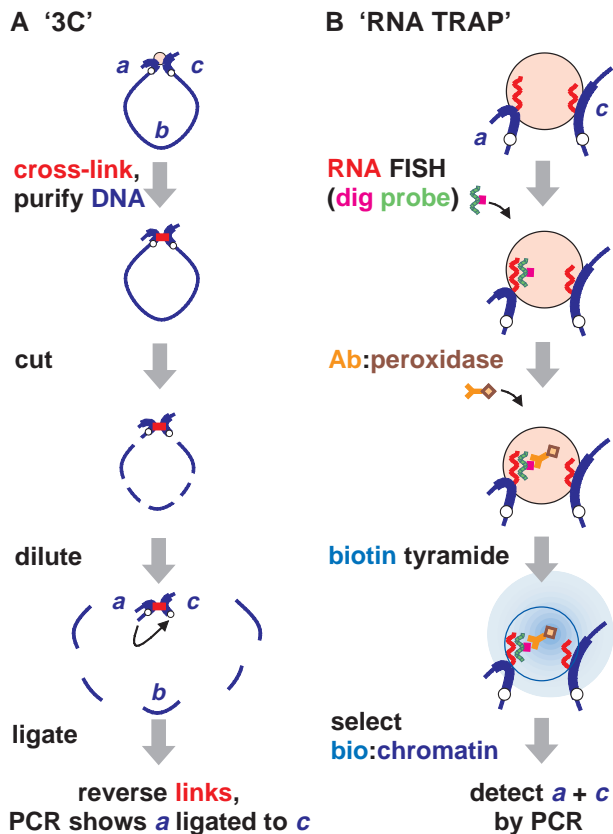
Supercoiling in linear eukaryotic DNA provides additional evidence for looping. Supercoils cannot be maintained in linear DNA without looping. However, lysing cells in >1 M NaCl releases superhelical loops that are visible in the electron microscope, and nascent transcripts are associated with attachment points (but not loops) (Jackson et al., 1984). Here, artifactual aggregation could induce looping, and so this is also indecisive. Many experiments involving nuclease digestion are also consistent with looping. Cutting an unlooped fibre should release long fragments that are then shortened, but the expected long fragments are not seen; rather, kinetics fit the release of short fragments from loops. Although initial experiments used unphysiological buffers that generated/destroyed attachments (Jackson et al., 1990), more recent ones with 'physiological' buffers confirm looping and show that transcription units mediate attachment (Fig. 2B) (Jackson et al., 1990; Jackson and Cook, 1993; Jackson et al., 1996).

Two powerful new methods now provide excellent evidence for looping (Fig. 3) and for a model described below in which the loops are anchored by transcription units. 'Chromosome conformation capture' (3C) (Dekker et al., 2002) shows that a distant (transcribed) LCR contacts the (transcribed) β-globin gene in mouse erythroid cells but not in brain cells (where the gene is inactive) (Tolhuis et al., 2002). Moreover, (transcribed) *scs* contacts (transcribed) *scs'*, and looping was confirmed by chromatin immunoprecipitation – antibodies against Zw5 (bound to *scs*) precipitate *scs'* (Blanton et al., 2003). The β-globin loop is also seen by RNA tagging and recovery of associated proteins (RNA TRAP) (Carter et al., 2002).

Although there is no decisive evidence for looping in living cells, the above data can be interpreted simply if two distant and active transcription units come together to tie the intervening fibre into a loop.

Combining models for loop structure and action at a distance

Models for action at a distance often involve looping (Merika and Thanos, 2001; Labrador and Corces, 2002; West et al., 2002; Schedl and Broach, 2003). Thus, a polymerizing complex might be attracted to an 'enhanceosome', before initiating to track out into a loop. But then transcription units should be detached in the experiment illustrated in Fig. 2 (but they are not), and they should not usually be seen together in

A '3C'

a ⚔ c
b

cross-link,
purify DNA

cut

dilute

a ⚔ c
b

ligate

reverse links,
PCR shows a ligated to c

B 'RNA TRAP'

a c

RNA FISH
(dig probe)

Ab:peroxidase

biotin tyramide

**select**
bio:chromatin

detect a + c
by PCR

Fig. 3. Two methods used to illustrate looping. (A) 3C. A loop is shown, tethered by transcription units *a* and *c* to a protein cluster (small pink circle). The 3C method involves fixation to covalently cross-link DNA sequences lying next to each other (usually through DNA:protein:DNA links shown as the red line), before removing unlinked proteins, cutting DNA, dilution and ligation. Dilution favours intramolecular ligation, so the ends of two different DNA molecules in one DNA:protein:DNA complex are joined together more frequently than those in different molecules/complexes. Then, two DNA sequences initially lying together in 3D space are often ligated together (i.e. *a* with *c*, but not *b*), and (after reversing cross-links) their juxtaposition is detected using PCR. (B) RNA TRAP. Only the region around transcription units *a* and *c* is illustrated. This method also involves fixation to preserve molecular contacts, in situ hybridization of a tagged intron probe to specific primary transcripts at a target transcription site (i.e. *a*), immunolabelling to bring peroxidase activity to the tagged probe, and reaction with biotin-tyramide to generate free-radicals that biotinylate proteins in the vicinity. Then DNA sequences initially lying near the target site in 3D space (*c*, but not *b*) co-purify with biotinylated chromatin and can be detected by PCR.

Territory
in nucleus

short
euchromatic
loop

long
hetero-
chromatic
loop

Cloud
around
factory

Loop
attached
to factory

Zig-zagging string

Nucleosome

DNA duplex

Fig. 4. Model for the organization of the chromatin fibre in a HeLa nucleus. DNA is wound into a nucleosome, and then a zig-zagging string of nucleosomes is tied to a factory through a cluster of transcription factors (diamond) or an active polymerase (oval). Components of the factory exchange with the soluble pool, and attachments to the factory are made and broken as factors dissociate and transcription terminates. 10-20 loops (only three are shown) of 5-200 kbp form a cloud around the transcription factory; long, static, loops are likely to become heterochromatic and attached to the lamina. 50-100 clouds then form a chromosome territory. Modified, with permission, from Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc. (Cook, 2001).

the experiments illustrated in Fig. 3 (but they are). Moreover, a tracking polymerase generates a transcript that is entwined about the template once for every ten base pairs transcribed, but no satisfactory method for untwining the transcript has yet been suggested (Cook, 1999). However, all the above results fit comfortably with one current model for chromatin structure (Cook, 1999; Cook, 2002) in which transcription complexes strung along the chromosome cluster to form a 'factory', and this would loop intervening DNA (Fig. 4). Here, each active polymerase in the factory reels in its template and extrudes its

transcript (Fig. 5). Because the template rotates as it moves through an immobile polymerase, the transcript does not become entwined about the template (Iborra et al., 1996; Cook, 1999). Support for this model comes from several approaches. First, active polymerases (plus their templates and transcripts) resist detachment in the experiment illustrated in Fig. 2B, which places them at (or very close to) attachment points (Jackson and Cook, 1993; Jackson et al., 1996; Cook, 1999). Second, because there are more active molecules of RNA polymerase II in a HeLa cell than transcription sites, and because only one polymerase is typically engaged on a transcription unit, each site (diameter ~50 nm) must contain
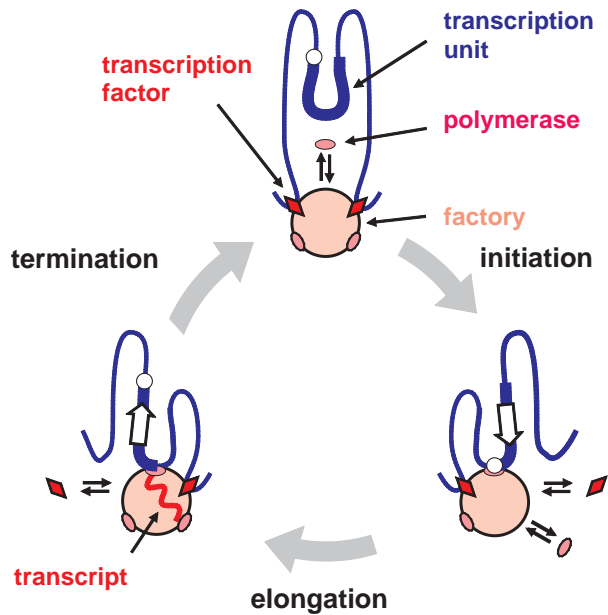
Fig. 5. A transcription cycle. A chromatin loop is attached through transcription factors to a factory; these attachments are made and broken continuously as components in the factory exchange with the soluble pool. On initiation, the promoter (white circle) binds to a polymerase in the factory, and the transcript is extruded as the template slides (white arrows) through the polymerase; on termination, the template detaches. Between initiation and termination the template is bound to the factory both stably (the bond is one of the stablest noncovalent ones known) and persistently (it takes ~5 minutes to transcribe a typical human gene) (Kimura et al., 2002). Therefore, such attachments are much more stable and persistent than those mediated by transcription factors, and this may underlie the difference in 'strength' between transcribed and nontranscribed enhancers.

several different units (Cook, 1999). The new results using 3C and RNA TRAP (Fig. 3) provide further support for the core element of this model – that two (or more) active units cluster to loop the intervening DNA; thus, the (transcribed) LCR contacts (transcribed) β-*globin* (Carter et al., 2002; Tolhuis et al., 2002), and (transcribed) *scs* contacts (transcribed) *scs'* (Blanton et al., 2003).

In this model, a distant regulator (a polymerizing complex) can easily contact the polymerase it regulates. Regulators would have to associate with the appropriate factories before they could exert their effects, because different factories specialize in transcribing different genes. Thus, some contain only RNA polymerase I, others contain only RNA polymerase II, and still others contain only RNA polymerase III (Pombo et al., 1999). Factories containing one kind of polymerase specialize even further. For example, a subset of polymerase II factories are rich in PSE-binding transcription factor (PTF) and Oct1, and they associate with particular chromosomes – presumably those carrying transcription units regulated by these factors (Pombo et al., 1998). In addition, genes encoding the histones – and U1, U2 and U3 snRNA – often lie near Cajal bodies (Spector, 2003), again presumably associated with factories dedicated to their transcription. It may even be that an individual factory could become dedicated to the transcription of one gene plus its flanking nongenic transcription units (for example, the β-globin gene, its LCR, plus other nongenic units). Only three types of 'functional' tie currently need to be invoked, although more may have to be added. These would involve (1) nontranscribed motifs (promoters, and a subset of what are currently called enhancers and silencers) bound to their target proteins in a factory, (2) transcribing DNA (including genic introns and exons, silencers, insulators, barriers, and a subset of what are currently called enhancers/LCRs) bound to active polymerases in a factory, and (3) appropriately modified nucleosomes out in the body (or bight) of a loop to other complementary nucleosomes in heterochromatin and/or the lamina (Polioudaki et al., 2001). The first would be transient, the second would persist for as long as the unit was transcribed (which is in the order of minutes) (Kimura et al., 2002) and the third could be almost permanent (because histones H3 and H4 in deep heterochromatin only exchange over a period of many hours) (Kimura et al., 2001).

This organization has important consequences (Cook, 2002). In the nucleus of a HeLa cell the concentration of soluble RNA polymerase II is ~1 µM, but the local concentration in a factory is 1000-fold higher. Because a promoter can diffuse ~100 nm in 15 seconds, one lying near a factory is likely to initiate; moreover, when released at termination it will still lie near a factory, and the movement and modifications (e.g. acetylation) accompanying elongation will leave it in an 'open' conformation. Another promoter out in a long loop (i.e. in the bight) is less likely to initiate because the promoter concentration falls off with the cube of the distance from the factory. Moreover, a long tether will buffer it from transcription-induced movement, making it prone to deacetylation, deposition of HP1 and incorporation into heterochromatin (Fig. 4). And because heterochromatin has an affinity for both the lamina (Polioudaki et al., 2001) and other heterochromatic regions (e.g. flanking ribosomal cistrons), inactive genes will inevitably be drawn to the periphery or to nucleoli (Cockell and Gasser, 1999; Galy et al., 2000; Spector, 2003). The context around a promoter will then be self-sustaining: productive collisions of an active promoter with the factory will attract factors increasing the frequency of initiation, and the longer an inactive promoter remains inactive the more it becomes embedded in heterochromatin.

The probability that promoters collide productively with a factory is increased by increasing promoter mobility (by 'opening' chromatin), increasing promoter-factory affinity (through binding of appropriate factors; Fig. 6A,B) and reducing promoter-factory distance (by shortening the tether; Fig. 6C,D) (Iborra et al., 1996). However, transcriptional interference will occur if the tether becomes too short (Fig. 6E). Some active transcription units will also be barriers separating euchromatin from heterochromatin (Fig. 6F). A promoter embedded in a long heterochromatic loop then becomes active by progressively activating units deeper and deeper into the loop; this 'opens' chromatin, subdivides the loop into ever smaller ones and brings the promoter closer to the factory. Only then can it compete effectively with others in the vicinity for polymerases in the factory. Here, binary switches do not activate genes flanked by discrete barriers; rather, a surrounding pattern of transcription encodes a fuzzy logic that specifies the probability of initiating.

Fig. 6. General (left) and specific views (right) of how transcriptional activity is regulated by local (A,B) and distant motifs (C-F). Left: thick and thin blue lines represent transcription units (with promoters as white circles) and intervening DNA, respectively. Right: half a loop is shown with unit *t* temporarily attached to, and transcribed by, a polymerase in a factory (pink circle); as a transcript is extruded (wavy red line), chromatin between *s* and *t* is reeled in and 'opened'. (A,B) Left: a repressor (or activator) binds close to the transcription start site to reduce (increase) the chances that a polymerase will be recruited. Right: repressor (or activator) binding to the promoter of *s* reduces (increases) the chances that *s* binds to the factory. In both cases, the repressor (or activator) has the same effect by binding to the polymerase (rather than the promoter). (C) Left: inserting a nontranscribed motif (e.g. some enhancers) increases the activity of a distant unit. Right: the motif has an affinity for transcription factors in the factory, and – once it binds – promoter *s* becomes tethered close to the factory, increasing its chances of attaching and initiating. A nontranscribed silencer (not shown) could have the opposite effect and reduce the chances that promoter *s* could bind (perhaps by directly blocking access to polymerases in the factory, or indirectly because DNA between the motif and promoter *s* was too short/rigid to loop back). (D) Left: inserting a transcribed motif (e.g. some enhancers, LCRs) increases the activity of a distant unit. Right: unit *s* attaches to the factory and initiates; this brings promoter *r* closer to the factory, increasing its chances of attaching and initiating. If the orientation of *s* were reversed, its transcription would progressively tether *r* ever closer to the factory and this might explain why orientation affects the activity of the β-globin LCR (Tanimoto et al., 1999). (E) Left: inserting an active unit near another, silences one of the two. Right: the attachment of unit *t* prevents *s* from attaching, perhaps because all polymerases are now occupied, DNA between *t* and *s* is too short/rigid to loop back, or transcription of *t* leads to recruitment of inappropriate factors to the factory; alternatively, *r* might attach to another (blue) factory lacking the polymerase and/or factors required by *s*, so distancing *s* from the (pink) factory with the appropriate polymerase/factors (*tRNA^Thr* in *HMR* probably silences *URA3* like this). (F) Left: Heterochromatin spreads leftwards (grey arrow) down the fibre, but inserting a barrier (yellow) restricts the spread and allows the unit to be expressed. Right: unit *t* acts as a barrier preventing heterochromatic spread to inactivate *s*.
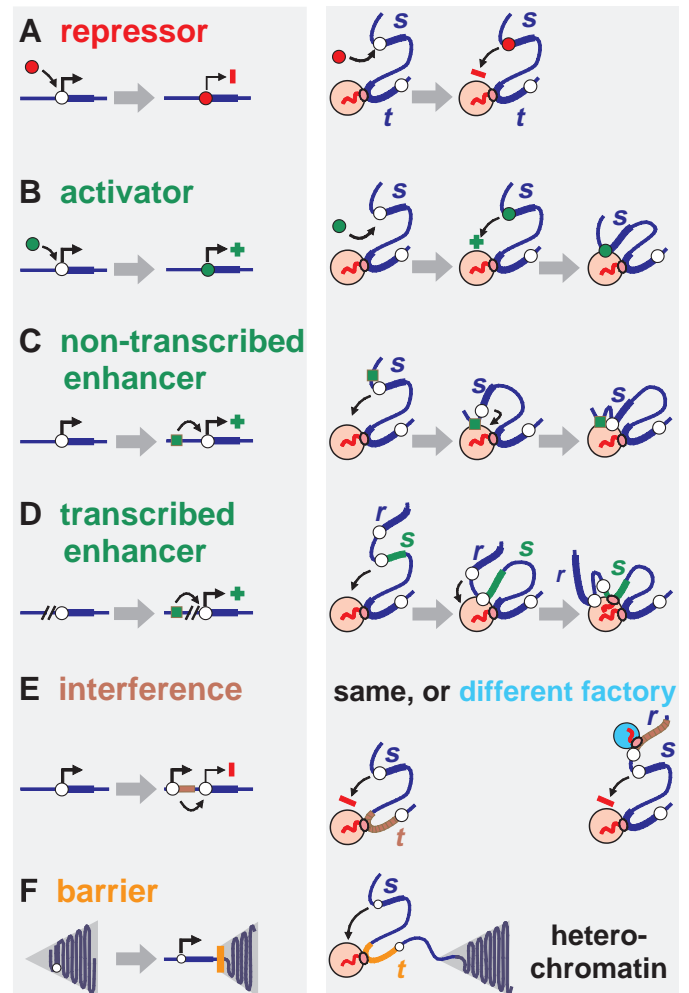
Conclusions

Many models invoke looping to explain how the chromatin fibre is organized during interphase and how regulatory motifs act at a distance. However, the two sets of models usually involve different molecular ties, largely because cell biologists working on different nuclear isolates have identified different ties – and then molecular biologists distrust the conflicting models they suggest. The model described here integrates aspects of both sets. It is based on the remarkable facts that clusters of active polymerases seem both to mediate action at a distance and to constitute the ties. It has several advantages. First, it is general, applicable to the range of motifs discussed and from yeast to man (although distance from the factory and nongenic transcription play larger roles as genome size increases). Second, it is consistent with current views on how everchanging interactions underlie self-assembling, but persistent, architectures (Misteli, 2001). Third, it suggests new possibilities. For example, it is difficult to explain why some eukaryotic genes and their regulatory motifs are so long (e.g. in BX-C). But, if this model applies, the activation of a long unit will tether the promoter close to a factory for longer, and this gives more time for that promoter to reinitiate; then, a longer elongation time is offset by a shorter time between initiations. Similarly, RNA polymerase II transcribes well beyond the poly(A) signal, but this apparently irrelevant transcription will nevertheless continue to tether the promoter close to the factory and so increase the rate of re-initiation (within any constraints imposed by transcriptional interference); polymerases paused anywhere in a transcription unit will have similar effects. Fourth, it highlights the shortcomings of current assays being used to analyze regulatory motifs, and helps to explain why the results obtained have been so difficult to interpret. These assays usually involve moving a test sequence to a different genomic region where a different (and usually unknown) set of flanking transcription units will inevitably exert their effects. In future, the first step in this kind of analysis should be the characterization of all transcription units in both the original and final locations. By the same token, analysis of the regulation of a gene should begin with the systematic characterization of all the flanking transcription units. Fifth, the model is testable. Thus, (1) most regulatory motifs should be transcription units, (2) point mutations knocking out their promoters should abolish their activity, and (3) adding/deleting promoters should affect expression of neighbouring genes.

The challenge now is to predict whether a gene is likely to be active from the DNA sequence. Unfortunately, we are still a long way from being able to do so in higher eukaryotes. At the very least, we need to know the location of all flanking transcription units, their relative activities, the concentration of the required transcription factors, how the length of the tether affects initiation rates and how heterochromatinization reduces DNA mobility. Only then can we predict which factories might be in the vicinity, and how the complex balance of conflicting forces that we know as context might be resolved.

References

Abranches, R., Beven, A. F., Aragon-Alcaide, L. and Shaw, P. J. (1998). Transcription sites are not correlated with chromosome territories in wheat nuclei. *J. Cell Biol.* **143**, 5-12.

Adachi, N. and Lieber, M. R. (2002). Bidirectional gene organization: a common architectural feature of the human genome. *Cell* **109**, 807-809.

Ashe, H. L., Monks, J., Wijgerde, M., Fraser, P. and Proudfoot, N. J. (1997). Intergenic transcription and transinduction of the human beta-globin locus. *Genes Dev.* **11**, 2494-2509.

Avramova, Z. and Tikhonov, A. (1999). Are scs and scs' 'neutral' chromatin domain boundaries of the locus? *Trends Genet.* **15**, 138-139.

Baker, W. K. (1968). Position-effect variegation. *Adv. Genetics* **14**, 133-169.

Belmont, A. S. (2002). Mitotic chromosome scaffold structure: new approaches to an old controversy. *Proc. Natl. Acad. Sci. USA* **99**, 15855-15857.

Bender, W. and Fitzgerald, D. P. (2002). Transcription activates repressed domains in the Drosophila bithorax complex. *Development* **129**, 4923-4930.

Bi, X. and Broach, J. R. (1999). UASrpg can function as a heterochromatin boundary element in yeast. *Genes Dev.* **13**, 1089-1101.

Blanton, J., Gaszner, M. and Schedl, P. (2003). Protein:protein interactions and the pairing of boundary elements in vivo. *Genes Dev.* **17**, 664-675.

Bolton, E. C. and Boeke, J. D. (2003). Transcriptional interactions between yeast tRNA genes, flanking genes and Ty elements: a genomic point of view. *Genome Res.* **13**, 254-263.

Burgess-Beusse, B., Farrell, C., Gaszner, M., Litt, M., Mutskov, V., Recillas-Targa, F., Simpson, M., West, A. and Felsenfeld, G. (2002). The insulation of genes from external enhancers and silencing chromatin. *Proc. Natl. Acad. Sci. USA* **99 (Suppl. 4)**, 16433-16437.

Carter, D., Chakalova, L., Osborne, C. S., Dai, Y.-f. and Fraser, P. (2002). Long-range chromatin regulatory interactions in vivo. *Nat. Genet.* **32**, 623-626.

Cheng, T. H. and Gartenberg, M. R. (2000). Yeast heterochromatin is a dynamic structure that requires silencers continuously. *Genes Dev.* **14**, 452-463.

Cockell, M. and Gasser, S. M. (1999). Nuclear compartments and gene regulation. *Curr. Opin. Genet. Dev.* **9**, 199-205.

Cook, P. R. (1999). The organization of replication and transcription. *Science* **284**, 1790-1795.

Cook, P. R. (2001). *Principles of nuclear structure and function*. New York: J. Wiley.

Cook, P. R. (2002). Predicting three-dimensional genome structure from transcriptional activity. *Nat. Genet.* **32**, 347-352.

Dekker, J., Rippe, K., Dekker, M. and Kleckner, N. (2002). Capturing chromosome conformation. *Science* **295**, 1306-1311.

Donze, D. and Kamakaka, R. T. (2001). RNA polymerase III and RNA polymerase II promoter complexes are heterochromatin barriers in Saccharomyces cerevisiae. *EMBO J.* **20**, 520-531.

Donze, D. and Kamakaka, R. T. (2002). Braking the silence: how heterochromatic gene repression is stopped in its tracks. *Bioessays* **24**, 344-349.

Drewell, R. A., Bae, E., Burr, J. and Lewis, E. B. (2002a). Transcription defines the embryonic domains of cis-regulatory activity at the Drosophila bithorax complex. *Proc. Natl. Acad. Sci. USA* **99**, 16853-16858.

Drewell, R. A., Arney, K. L., Arima, T., Barton, S. C., Brenton, J. D. and Surani, M. A. (2002b). Novel conserved elements upstream of the H19 gene are transcribed and act as mesodermal enhancers. *Development* **129**, 1205-1213.

Emerman, M. and Temin, H. M. (1986). Quantitative analysis of gene suppression in integrated virus vectors. *Mol. Cell. Biol.* **6**, 792-800.

Galy, V., Olivo-Marin, J. C., Scherthan, H., Doye, V., Rascalou, N. and Nehrbass, U. (2000). Nuclear pore complexes in the organization of silent telomeric chromatin. *Nature* **403**, 108-112.

Gerasimova, T. I. and Corces, V. G. (2001). Chromatin insulators and boundaries: effects on transcription and nuclear organization. *Annu. Rev. Genet.* **35**, 193-208.

Gottesman, S. (2002). Stealth regulation: biological circuits with small RNA switches. *Genes Dev.* **16**, 2829-2842.

Grunstein, M. (1990). Histone function in transcription. *Annu. Rev. Cell Biol.* **6**, 643-678.

Hertel, K. J., Lynch, K. W. and Maniatis, T. (1997). Common themes in the function of transcription and splicing enhancers. *Curr. Opin. Cell Biol.* **9**, 350-357.

Hirotsune, S., Yoshida, N., Chen, A., Garrett, L., Sugiyama, F., Takahashi, S., Yagami, K., Wynshaw-Boris, A. and Yoshiki, A. (2003). An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. *Nature* **423**, 91-96.

Hogga, I. and Karch, F. (2002). Transcription through the iab-7 cis-regulatory domain of the bithorax complex interferes with maintenance of Polycomb-mediated silencing. *Development* **129**, 4915-4922.

Iborra, F. J., Pombo, A., McManus, J., Jackson, D. A. and Cook, P. R. (1996). The topology of transcription by immobilized polymerases. *Exp. Cell Res.* **229**, 167-173.

Iborra, F. J., Jackson, D. A. and Cook, P. R. (2001). Coupled transcription and translation within nuclei of mammalian cells. *Science* **293**, 1139-1142.

Ivarie, R. D., Schacter, B. S. and O'Farrell, P. H. (1983). The level of expression of the rat growth hormone gene in liver tumour cells is at least eight orders of magnitude less than that in anterior pituitary cells. *Mol. Cell. Biol.* **3**, 1460-1467.

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

Jackson, D. A., McCready, S. J. and Cook, P. R. (1984). Replication and transcription depend on attachment of DNA to the nuclear cage. *J. Cell Sci.* **(Suppl. 1)**, 59-79.

Jackson, D. A., Dickinson, P. and Cook, P. R. (1990). The size of chromatin loops in HeLa cells. *EMBO J.* **9**, 567-571.

Jackson, D. A., Bartlett, J. and Cook, P. R. (1996). 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.

Jenuwein, T. and Allis, C. D. (2001). Translating the histone code. *Science* **293**, 1074-1080.

Johnson, K. D., Christensen, H. M., Zhao, B. and Bresnick, E. H. (2001). Distinct mechanisms control RNA polymerase II recruitment to a tissue-specific locus control region and a downstream promoter. *Mol. Cell* **8**, 465-471.

Kapranov, P., Cawley, S. E., Drenkow, J., Bekiranov, S., Strausberg, R. L., Fodor, S. P. and Gingeras, T. R. (2002). Large-scale transcriptional activity in chromosomes 21 and 22. *Science* **296**, 916-919.

Kellum, R. and Schedl, P. (1991). A position-effect assay for boundaries of higher order chromosomal domains. *Cell* **64**, 941-950.

Kimura, H. and Cook, P. R. (2001). Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. *J. Cell Biol.* **153**, 1341-1353.

Kimura, H., Sugaya, K. and Cook, P. R. (2002). The transcription cycle of RNA polymerase II in living cells. *J. Cell Biol.* **159**, 777-782.

Kurdistani, S. K. and Grunstein, M. (2003). Histone acetylation and deacetylation in yeast. *Nat. Rev. Mol. Cell. Biol.* **4**, 276-284.

Labrador, M. and Corces, V. G. (2002). Setting the boundaries of chromatin domains and nuclear organization. *Cell* **111**, 151-154.

Li, Q., Peterson, K. R., Fang, X. and Stamatoyannopoulos, G. (2002). Locus control regions. *Blood* **100**, 3077-3086.

Lindsley, D. L., Edington, C. W. and von Halle, E. S. (1960). Sex-linked recessive lethals in *Drosophila* whose expression is suppressed by the Y-chromosome. *Genetics* **45**, 1649-1670.

Litt, M. D., Simpson, M., Gaszner, M., Allis, C. D. and Felsenfeld, G. (2001). Correlation between histone lysine methylation and developmental changes at the chicken beta-globin locus. *Science* **293**, 2453-2455.

Martin, C. H., Mayeda, C. A., Davis, C. A., Ericsson, C. L., Knafels, J. D.,

Mathog, D. R., Celniker, S. E., Lewis, E. B. and Palazzolo, M. J. (1995). Complete sequence of the bithorax complex of Drosophila. *Proc. Natl. Acad. Sci. USA* **92**, 8398-8402.

Merika, M. and Thanos, D. (2001). Enhanceosomes. *Curr. Opin. Genet. Dev.* **11**, 205-208.

Misteli, T. (2001). The concept of self-organization in cellular architecture. *J. Cell Biol.* **155**, 181-185.

Partridge, J. F., Borgstrom, B. and Allshire, R. C. (2000). Distinct protein interaction domains and protein spreading in a complex centromere. *Genes Dev.* **14**, 783-791.

Plath, K., Mlynarczyk-Evans, S., Nusinow, D. A. and Panning, B. (2002). Xist RNA and the mechanism of x chromosome inactivation. *Annu. Rev. Genet.* **36**, 233-278.

Polioudaki, H., Kourmouli, N., Drosou, V., Bakou, A., Theodoropoulos, P. A., Singh, P. B., Giannakouros, T. and Georgatos, S. D. (2001). Histones H3/H4 form a tight complex with the inner nuclear membrane protein LBR and heterochromatin protein 1. *EMBO Rep.* **2**, 920-925.

Pombo, A., Cuello, P., Schul, W., Yoon, J.-B., Roeder, R. G., Cook, P. R. and Murphy, S. (1998). 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.

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

Rank, G., Prestel, M. and Paro, R. (2002). Transcription through intergenic chromosomal memory elements of the Drosophila bithorax complex correlates with an epigenetic switch. *Mol. Cell. Biol.* **22**, 8026-8034.

Richards, E. J. and Elgin, S. C. (2002). Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. *Cell* **108**, 489-500.

Routledge, S. J. and Proudfoot, N. J. (2002). Definition of transcriptional promoters in the human beta globin locus control region. *J. Mol. Biol.* **323**, 601-611.

Schedl, P. and Broach, J. R. (2003). Making good neighbors: the right fence for the right job. *Nat. Struct. Biol.* **10**, 241-243.

Shendure, J. and Church, G. M. (2002). Computational discovery of sense-antisense transcription in the human and mouse genomes. *Genome Biol.* **3**, RESEARCH0044.

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P. J., Cordum, H. S., Hillier, L., Brown, L. G., Repping, S., Pyntikova, T., Ali, J., Bieri, T. et al. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, **423**, 825-837.

Snow, M. H. L. and Callan, H. G. (1969). Evidence for a polarized movement of the lateral loops of newt lampbrush chromosomes during oogenesis. *J. Cell Sci.* **5**, 1-25.

Spector, D. L. (2003). The dynamics of chromosome organization and gene regulation. *Annu. Rev. Biochem.* **72**, 573-608.

Storz, G. (2002). An expanding universe of noncoding RNAs. *Science* **296**, 1260-1263.

Sturtevant, A. H. and Beadle, G. W. (1962). *An introduction to Genetics.* New York: Dover Publications.

Sudbrak, R., Wieczorek, G., Nuber, U. A., Mann, W., Kirchner, R., Erdogan, F., Brown, C. J., Wohrle, D., Sterk, P., Kalscheuer, V. M. et al. (2001). X chromosome-specific cDNA arrays: identification of genes that escape from X-inactivation and other applications. *Hum. Mol. Genet.* **10**, 77-83.

Sutter, N. B., Scalzo, D., Fiering, S., Groudine, M. and Martin, D. L. (2003). Chromatin insulation by a transcriptional activator. *Proc. Natl. Acad. Sci. USA* **100**, 1105-1110.

Tanimoto, K., Liu, Q., Bungert, J. and Engel, J. D. (1999). Effects of altered gene order or orientation of the locus control region on human beta-globin gene expression in mice. *Nature* **398**, 344-348.

Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell.* **10**, 1453-1465.

Volpe, T. A., Kidner, C., Hall, I. M., Teng, G., Grewal, S. I. and Martienssen, R. A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**, 1833-1837.

West, A. G., Gaszner, M. and Felsenfeld, G. (2002). Insulators: many functions, many mechanisms. *Genes Dev.* **16**, 271-288.

Willoughby, D. A., Vilalta, A. and Oshima, R. G. (2000). An Alu element from the K18 gene confers position-independent expression in transgenic mice. *J. Biol. Chem.* **275**, 759-768.

Wyrick, J. J., Holstege, F. C., Jennings, E. G., Causton, H. C., Shore, D., Grunstein, M., Lander, E. S. and Young, R. A. (1999). Chromosomal landscape of nucleosome-dependent gene expression and silencing in yeast. *Nature* **402**, 418-421.