

# Transcription-Based Solenoidal Model of Chromosomes

François Képès<sup>a,b</sup> Cédric Vaillant<sup>a</sup>

<sup>a</sup>Atelier de Génomique Cognitive, CNRS UMR8071/Génopole<sup>®</sup> and <sup>b</sup>Epigenomics Project, Génopole<sup>®</sup>, Evry, France

## Key Words

Transcription · DNA structure · Nucleus, functional organization · DNA-binding proteins · Yeast · Bacteria

## Abstract

The organization of transcription within the eukaryotic nucleus or the prokaryotic nucleoid may be expected to both depend on and determine the structure of the chromosomes. In yeast and bacteria, genes that are controlled by the same sequence-specific transcription factor tend to be either clustered or regularly spaced along the chromosomes. The same spacing is found for most transcription factors within a chromosome. Furthermore, in bacteria, the gene encoding the transcription factor tends to locate at identical regular intervals from its targets. This periodicity is consistent with a solenoidal epi-organization of the chromosome, which would dynamically gather the interacting partners into foci. Binding at genuine regulatory sites on DNA would thus be optimized by locally increasing the concentration of transcription factors and their binding sites. As many transcription factors are simultaneously active and some share targets, the resulting collection of foci provides a potent self-organizational principle for the chromosome, and consequently for the functional nuclear architecture.

Copyright © 2003 S. Karger AG, Basel

Fax +41 61 306 12 34  
E-Mail karger@karger.ch  
www.karger.com

© 2003 S. Karger AG, Basel  
1424–8492/03/0014–0171  
\$19.50/0

Accessible online at:  
www.karger.com/cpu

François Képès, ATGC, CNRS UMR8071/Génopole<sup>®</sup>  
523 Terrasses de l'Agora  
FR–91000 Evry (France)  
Tel. +33 1 60 87 38 02, Fax +33 1 60 87 38 09,  
E-Mail Francois.Kepes@genopole.cnrs.fr

## Synopsis

Biologists now have access to a virtually complete map of all the genes in the human genome. Researchers are aggressively assembling a similarly detailed knowledge of the 'proteome' – the full collection of proteins that are encoded by those genes. And they are going further, probing the 'transcriptome' – the diverse set of mRNA molecules that link genes to proteins and serve as templates for protein manufacture. In broad outline, the fundamentals of molecular biology are simple: information flows from genes to mRNAs (through the process of 'transcription') and then from mRNAs to proteins (through 'translation'). Yet this picture leaves out an immense backdrop of complex feedback, and there remains a deep mystery: how is the overall process of gene expression controlled? How are genes turned on and off with such delicate and adaptive precision?

Even the most basic processes continue to turn up surprises. Until a few years ago, for example, biologists believed that mRNA molecules – once 'read out' from the genes – almost always went on to direct the manufacture of specific proteins. No one suspected that regulatory dynamics might operate well downstream from transcription, based on so-called 'RNA interference'. Researchers now understand that the intermediary RNA molecules can be marked out for destruction, effectively 'silencing' the expression of their associated gene.

The regulation of genes is also influenced by other factors, such as physical geometry. Within our chromosomes, the double-stranded DNA molecule wraps tightly around a range of molecules called histones, forming a beaded structure, which then twists into a secondary helix on a larger scale. The resulting 'chromatin fiber' is then packed into yet more complex structures at higher levels to yield the basic material of the chromosomes, a material that evolves dynamically, becoming more or less compact depending on the state of the cell. In so doing, it allows regulatory

**KARGER**

## Introduction

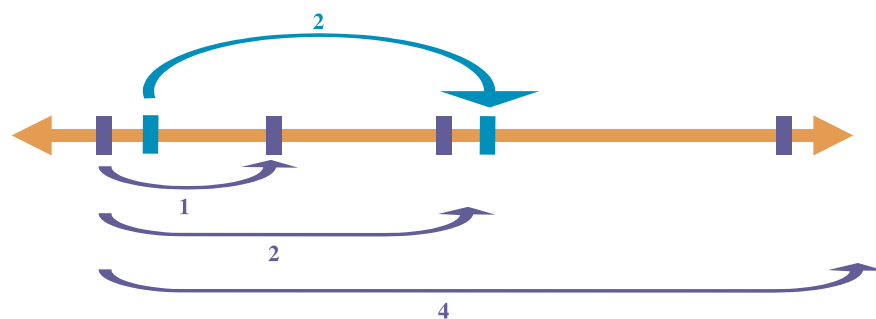
Predicting the three-dimensional structure of chromosomes from the primary DNA sequence has become an important goal, as genomic and transcriptomic data are now generated at an elevated pace. The transcription-based solenoidal model of chromosomes represents a step towards this goal. This model connects transcriptomic data to the functional and spatial organization of chromosomes. This paper outlines the model and describes some of its supporting evidence and consequences.

## Observations

Individual gene transcription is modulated by sequence-specific transcription factors (TF). A TF is a protein that binds to a specific DNA site in the regulatory region of its target gene to activate or repress its transcription. Both in the eukaryotic baker's yeast *Saccharomyces cerevisiae* [1] and in the prokaryotic enterobacterium *Escherichia coli* [2], target genes that are controlled by the same TF tend to be either clustered or regularly spaced along the chromosome. The same spacing is observed for most TFs within the *E. coli* cir-

cular chromosome in the nucleoid or within any of the 32 yeast chromosome arms in the nucleus. Importantly, the periods differ among the yeast chromosome arms and among four different *E. coli* strains. Furthermore, in *E. coli*, target genes locate at periodic intervals from the gene encoding their TF. The TF/target and target/target regular periods are both 1/50th of the chromosome circumference. This TF/target pattern is more pronounced than the target/target one, thereby suggesting that the former causes the latter [2]. In *S. cerevisiae*, no specific location of the TF's gene with respect to its target genes is observed. However, replication initiation sites are sometimes found to locate at regular intervals equal to the transcriptional period of the same chromosome arm [1].

These findings relied on the analysis of datasets that relate each TF to a list of its targets determined by bench experimentation, not by sequence-based prediction. Either the list was mined from the literature [3,4] or it was obtained with a recent high-throughput method called chromatin immunoprecipitation [5,6]. The data analysis is depicted on figure 1.



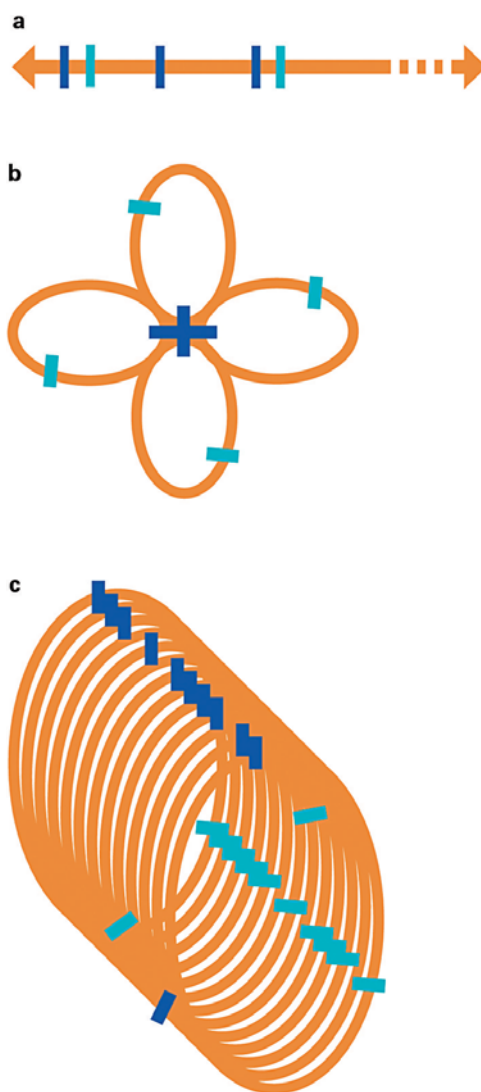
**Fig. 1.** Analysis of the distance between regularly spaced genes. The distances are multiples (denoted by numbers) of one period, corresponding to a fixed number of base pairs. Dark rectangles represent targets of one TF, light rectangles represent targets of a second TF. For the sake of clarity, only the distances from the first occurrence of each target type are recorded here.

molecules to access the genes, or inhibits access instead. The three-dimensional structure of DNA packing in chromosomes has a great deal to do with genetic regulation, even at the initial level of gene transcription.

Yet many of the geometric details of chromosome structure remain unknown. And in the present paper, Képès and Vaillant take a step toward filling in the picture. They examine the character of chromatin – the chromosomal material of packed DNA – at an intermediary level, and offer evidence to suggest that a specific ‘solenoidal’ topology of the packed DNA plays an important role in the regulation of transcription. Transcription factors are proteins that bind to specific sites along the DNA to either activate or repress the expression of certain genes. Earlier studies in two organisms – the yeast *Saccharomyces cerevisiae* and the bacterium *Escherichia coli* – have shown that the target genes regulated by a transcription factor tend to be spaced in a regular, periodic way along the DNA with periods ranging from 10–100 genes (or roughly 20–100 kbp). This striking periodicity would seem like more than an accident; Képès and Vaillant believe that it has structural and regulatory significance.

Specifically, they propose that this periodicity reflects a higher-level coiling of the chromosomal DNA that brings many co-regulated genes together into one location, providing easy and efficient access for their associated transcription factors. They further argue that this coiling may be caused by the very dynamics that it allows. The coexistence of several target genes in one physical location should act to bring transcription factor molecules into that region. In turn, a higher density of transcription factors should recruit more target genes. Képès and Vaillant suggest that this transcription-based nonlinear feedback – driven by the kinetics of sequence-specific binding between transcription factors and their target genes – may be the force that

**Fig. 2.** Various arrangements of the DNA fiber. Dark rectangles represent targets of one TF, light rectangles represent targets of a second TF. **a** Linear. None of the TF targets are spatially clustered. **b** Radial loops or rosettes. The dark targets are clustered, but the light ones cannot be grouped. **c** Solenoidal loops. Both target sets are clustered, even though a few individual targets may not be at the proper position.



### Transcription-Based Solenoidal Model of Chromosomes

Based on these results, a solenoidal topology of chromosomes was proposed. Essentially, this model posits that the interacting partners, i.e. several copies of a given TF and of its DNA binding sites, gather within the nucleus/nucleoid into small subvolumes called ‘foci’. A TF may sometimes be found in a small number of foci that each contain a different group of target genes. Conversely, one focus may sometimes contain a few TFs of different types that, together, regulate common target genes.

Spatial confinement of target genes, each with several binding sites, synergistically recruits more of their TF. In turn, elevated concentration of the generally bivalent TF synergistically recruits more binding sites. This self-stimulating loop eventually increases the local concentrations of both TFs and binding sites. A similar mechanism is at play in immunoprecipitations, where the antigen is polyvalent with respect to the polyclonal antibody, and the antibody is bivalent. Here, the polyvalent antigen corresponds to the polyvalent (several binding sites) target gene, and the bivalent antibody corresponds to the bivalent TF. Immunoprecipitation results

drives the DNA to fold into a roughly solenoidal topology, thereby making transcription and genetic control more efficient.

In the paper, the authors describe this elegant hypothesis and present some of the evidence in its favor. They also examine a number of puzzles that the idea would naturally explain, and a series of definitive predictions that might be used to test the hypothesis in the near future. The manuscript offers the full detail; I summarize a few of the main points here.

### Evidence

As the authors argue, several pieces of evidence point to the solenoidal structure for DNA within the chromosomes and its origin in transcriptional dynamics:

For different chromosome arms in yeast, and in distinct strains of *E. coli*, the positioning of the target genes along the DNA reveals clear periodicities, not only for one transcription factor, but for many distinct factors (see their references 1 and 2). Moreover, within any chromosome arm or bacterial strain, the spacing of these genes is always the same. Given that the target genes for different transcription factors are highly interspersed along the DNA, the most likely way to account for these many periodic patterns – all having precisely the same period – would seem to be through a fundamental geometric explanation. A solenoidal coiling of the DNA is the simplest and most obvious geometric structure that serves the purpose.

In principle, one might suppose that this solenoidal structure could be enforced through the chemical action of some dedicated physical scaffolding about which the DNA might wind. Yet the target-gene periodicities are different on different yeast chromosome arms and in the distinct strains of *E. coli*, whereas one might expect a constant period if DNA were wrapped around a dedicated physical scaffolding. Hence, a more plausible origin for the solenoidal structure lies in a dynamic feedback between the transcription factors and the

from cross-linking many antigens and antibodies. By analogy, high cooperativity and elevated local concentration result from cross-linking many binding sites and TFs.

In both yeast and bacteria, this elevated local concentration favors TF binding at specific DNA sites, thereby sequestering the TF away from spurious DNA binding sites. Furthermore, in bacteria, a protein is synthesized near its encoding gene in the frequent case where its polymerization initiates from messengers that are still tethered to DNA by ongoing transcription [7]. When the bacterial protein is a TF, the model implies that this TF will be produced near its target genes, thereby reducing the search time between synthesis and binding at appropriate DNA sites.

One way to achieve this local concentration effect is to cluster these genes along the linear structure of the DNA: '1-dimensional (1-D) clustering'. However, the unbranched structure of DNA and the span of some genes pose limits on the number of genes that may be clustered by being contiguous. A second way to increase local concentration beyond the limits of linear contiguity is through 3-dimensional (3-D) spatial proximity: '3-D clustering'. We propose that the most parsimonious way to accommodate the hypothetical 3-D clustering and the observed regular spacing is a solenoidal model of the chromosome, where the regular period would correspond to one turn of the solenoid.

### Evidence for the Model

The transcriptomic evidence in favor of this solenoidal model is as follows [1, 2]:

(1) Regular spacing for one TF does not demonstrate 3-D spatial proximity in a solenoid. However, the finding of an identical period for most TFs along any given chromosome (arm) strongly constrains the interpretation, because their genes and their targets are interspersed throughout the whole chromosome. Only a generalized local model, with a solenoidal rather than

a radial or toroidal topology, can parsimoniously account for period invariance both for interspersed genes and for the whole chromosome (fig. 2). In other words, period invariance must reflect the propagation along the whole chromosome of a local constraint generated by gene interspersing. This constraint is strengthened by the high number of TFs that function simultaneously and must share the circumference of the solenoid seen in face view (fig. 3a).

When a TF locates in more than one focus, it should correspondingly be found at several positions along the circumference of the chromosomal solenoid seen in face view. Conversely, when a focus contains more than one TF species, these different TFs should form a group along the circumference, probably to simultaneously regulate common target genes.

(2) The alternative would be that period invariance reflects a dedicated scaffold holding DNA in a solenoidal configuration, much like histones constrain DNA structure at a smaller scale. In this case, one would expect a constant period among yeast chromosomes, unless the scaffolding molecules have chromosome-dependent properties or unless different chromosomes have a highly divergent content. Instead, the periods differ among chromosomes.

(3) Eukaryotic centromeres are poorly transcribed regions and would thus stop the propagation of the above-discussed local constraint if it were of a dynamic nature. Indeed, the period often changes dramatically as centromeres are crossed, although it remains constant over a whole chromosome arm spanning up to 50 periods (fig. 3b). Hence, this observation strongly supports the notion that periodicity has been shaped by the dynamics of transcription.

(4) Because the bacterial chromosome is circular, the regular spacing imposes that the period be an exact fraction of its full circumference. It appears to be the case in all four studied *E. coli* strains, as the ma-

genes they regulate, which would produce a structure having a periodicity matched to the genes in each case.

The transcriptional origin of the solenoidal structure is also implicated by the empirically observed disruption of gene periodicity around the centromere – the central region of a chromosome. The periodicity of the target genes changes at the centromere, being different on the arms to either side. Significantly, it is well known that transcriptional activity is suppressed in the centromere. Hence, the solenoidal structure is least pronounced at precisely the point where a transcription-based explanation would predict.

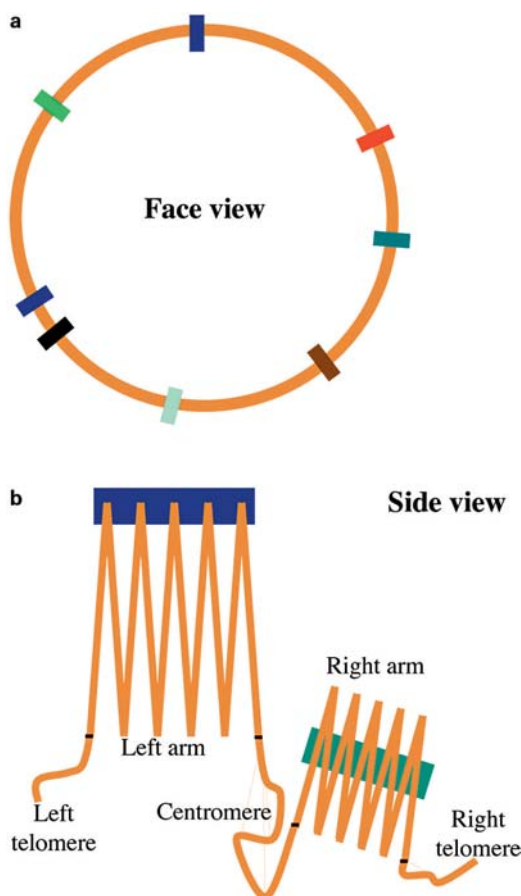
In the case of bacteria, the chromosomes are circular. Hence, the solenoidal hypothesis demands that the gene periodicity in this case must be a precise fraction of the overall chromosome length. As Képès and Vaillant note, this appears to be the case: 'in all four studied *E. coli* strains, the major period is exactly 1/50th of their chromosome circumference'.

### What the Hypothesis Explains

The solenoidal hypothesis offers simple explanations for a number of otherwise puzzling findings:

A major puzzle of gene transcription is that many sites along the DNA present transcription factors with excellent binding opportunities, yet these factors (proteins) do not in fact bind there. Transcription factors somehow manage to bind to their specific target sites even when other sites would allow stronger binding. Why? Képès and Vaillant note that an explanation follows readily from their hypothesis, at least for bacteria with no nuclear envelope in which transcription and translation are not physically separated. In this case, as they point out, the genes that encode transcription factors also follow the very same periodicity relationship as the target genes. Consequently, the transcription factors should be produced on the surface of the solenoid quite nearby the very

**Fig. 3.** Solenoidal model. **a** Face view of the topology of a prokaryotic chromosome or eukaryotic chromosome arm. Tens of TFs can be simultaneously active. Targets of a few of these numerous active TFs have been represented on this solenoid. Thus, one rectangle corresponds to several targets (superimposed in this face view) of the same TF. The circularity of the prokaryotic chromosome is not taken into account here. **b** Side view of the topology of a eukaryotic chromosome. Poorly transcribed regions such as centromere and telomeres do not propagate the local constraint imposed by the dynamics of transcription. Therefore, each chromosome arm is an independent solenoid. Indeed, in yeast, the right and left arms of a chromosome often have a different period, represented here by a different diameter. Each rectangle of panel **a** covers most turns of the solenoid, as shown here by one long rectangle. In essence, this rectangle depicts the volume where many copies of one TF and its binding site are locally concentrated. In the interest of clarity, only one rectangle per arm has been represented, but there should be one rectangle per active transcription cluster.



major period is exactly 1/50th of their chromosome circumference.

(5) 1-D clustering optimizes transcriptional initiation by focusing a TF and its multiple target DNA binding sites to increase their local concentrations [8,9]. For the sake of parsimony, a common mechanistic explanation is proposed here for both 1-D and 3-D clustering. Thus, to increase local concentration as does 1-D clustering, the periodically spaced sites must be spatially grouped, thereby achieving 3-D clustering.

### Additional Remarks

(1) To be imprinted on today's genomes, this optimizing principle must be evolvable and selectable, otherwise shuffling mechanisms such as recombination would have erased its traces. That it is selectable was already apparent from the studies on the trivalent lactose operator and its bivalent repressor [8, 9]. As spurious binding sites constantly arise through mutations, selective pressures have probably been quite strong to avoid unspecific binding by exploiting this local concentration effect. That it can evolve to the present situation under selective pressure is a reasonable assumption, as all it asks for is coevolution of

genes that are their targets. The solenoidal structure creates a physical proximity – not obvious from the DNA sequence itself – that can explain the puzzle of transcription factor specificity.

Another mystery of regulation is the remarkable speed with which transcription factors are able to find and act upon their target genes. If these proteins have to search the entire DNA strand, then regulation should take place slowly as the molecules take time to diffuse and find their specific targets. But this mystery vanishes if the transcription factors start out nearby their targets on the solenoidal surface. The solenoidal arrangement would help transcription factors to do their work rapidly.

On another general point, it is somewhat surprising, at least naively, that the regulation of genes takes place at all, given the dense packing of the DNA within the chromosome. This is less mysterious if DNA packing takes place precisely in a way that helps transcription factors to locate and bind to their targets efficiently.

### Predictions

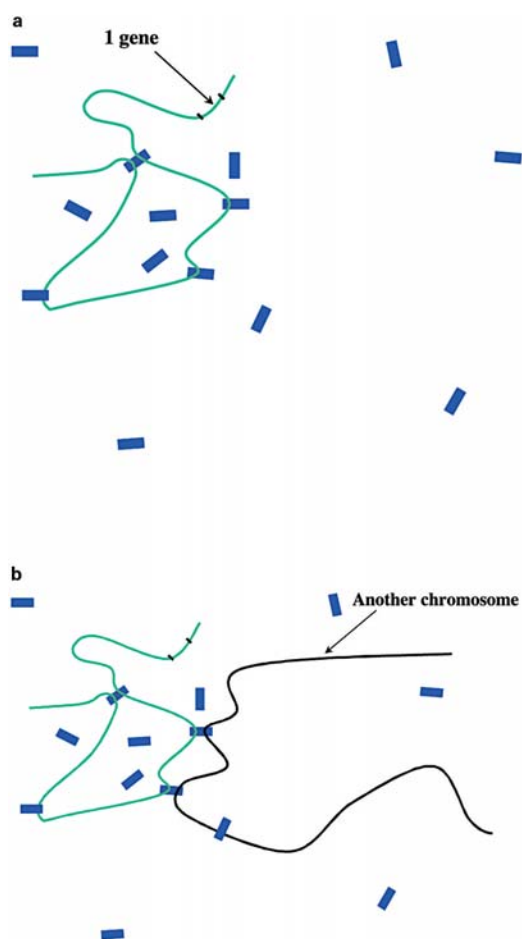
Finally, the solenoidal hypothesis naturally makes several predictions that might easily be tested in the future:

If the solenoidal structure plays a major role in enabling transcription, then one would expect the same kinds of periodicities to be found among the target genes of transcription factors in other organisms. Studies extended beyond *S. cerevisiae* and *E. coli* should find similar results.

If gene periodicity plays a key role in the regulation of transcription, then the departure of genes away from the preferred location should have specific consequences. In particular, transcriptional efficiency should vary with the location of target genes, with those located more precisely at the proper periodic position being expressed more strongly than those displaced further away.

Also, if the placement of target genes within the periodic scheme is so important

**Fig. 4.** Extending the concept of local concentration effect. Clustering ensures that the local concentration of both TFs and binding sites is high in the active areas, a self-sustaining and self-enhancing process. Line: Chromosome; rectangle: bivalent TF. **a** From intragenic to intergenic. The boundaries of one gene on the chromosome (light) are shown for scale. The bivalent TF bridges two remote segments of the chromosome. **b** From intrachromosomal to interchromosomal. A second chromosome (dark) is shown. Now some bivalent TFs bridge both chromosomes.



target positions with respect to each other. This multiple and simultaneous coevolutionary process has no reason to converge to any particular period, and this remark is compatible with the observed variety of periods.

(2) There are still some limitations to the periods that will eventually be reached [10]. For instance, periods cannot be too short due to the stiffness of the DNA-protein fiber (no less than 8–9 kilobase pairs (kbp) in eukaryotes [11]). Periods cannot be too long either, otherwise regulatory sites may not recur often enough to accommodate the transcriptional scheme. In *E.*

*coli* K12, the major period is 92.8 kbp, i.e. about 100 genes, and there are minor periods at 46.4 and 371.2 kbp (0.5- and 4-fold the major period, respectively). In *S. cerevisiae*, the periods vary among arms from about 9–50 kbp, with an average of 22 kbp, i.e. about 10 genes.

(3) The solenoidal model may be viewed as an extension of the local concentration concept from the intragenic [8,9] to the intergenic situation (fig. 4a), furthered in eukaryotes by an extension from the intrachromosomal to the interchromosomal situation (fig. 4b).

– as it places these genes in physical proximity with one another and with their transcription factors – then any disturbance of the periodic structure should have a detrimental effect on transcription. In particular, the random insertion or deletion of sections of ‘junk’ DNA from the DNA strand should disturb the precise periodic organization and have measurable consequences for transcription.

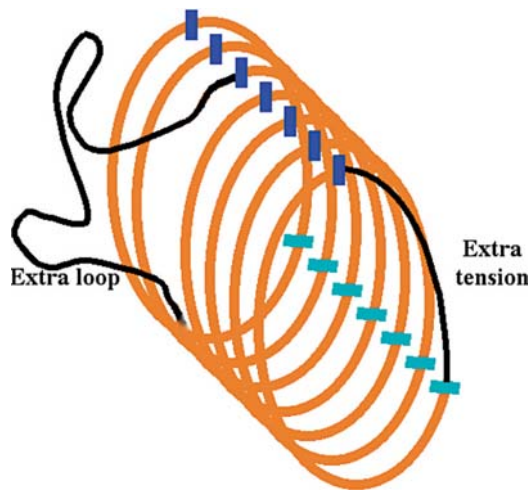
As the authors emphasize, their hypothesis suggests that DNA coils into a solenoidal ‘topology’ rather than a strict solenoidal ‘geometry’. There is no suggestion that the DNA falls into a perfect geometric tube. It is only necessary that the coiling brings the various target genes into the same neighborhood; hence, one should expect local distortions of the tube-like structure.

This hypothesis appears to offer an elegant account of a diverse range of observations, and to provide the motivation for a number of potentially illuminating experiments.

In addition, it may well provide an intuitive framework for understanding the great versatility of the cellular transcription machinery, which responds in a concerted and adaptive fashion to environmental changes. As the authors have suggested elsewhere, such adaptive power might arise if ‘alternative solenoids’ were imprinted on the same chromosome, thereby facilitating a rapid and tightly orchestrated transcriptional response to new environmental demands.

Mark Buchanan

**Fig. 5.** Imperfect solenoidal topology. Some arcs of the solenoid may be replaced by shorter (right side) or longer (left side) segments (in black). The high number of target clusters (only two are represented here) poses limits on such deviations from perfect topology. The data indicate that such deviations do not occur frequently.



(4) The solenoidal topology should not be understood as a perfect solenoidal geometry. The only requirement is that the many active transcriptional foci must coexist in various places of the nucleus/nucleoid, with the understanding that they all contain DNA sites that are equally spaced from each other, hence the solenoidal topology. As many TFs are simultaneously active and some share targets, the resulting collection of foci provides a potent organizational principle for the genome.

(5) One could imagine that, between two successive sites corresponding to two different TFs, it would be possible either to pull on the DNA rope or to push in a long extra DNA loop (fig. 5). While this is certainly possible, it would increase the background noise in the target/target or TF/target distance distributions. The results [1, 2] indicate that such phenomena, if any, must be limited.

### Circumstantial Evidence

A wide spectrum of data originating from unrelated studies provides circumstantial evidence for the solenoidal model.

As far as eukaryotes are concerned, the yeast results [1] support a correlation between chromosomal position, transcriptional activity and replication origin. This is reminiscent of the proposed correlation in mammalian nuclei between chromosomal territories, distinct transcriptional activity levels and replication timing [reviewed in 12]. This analogy is strengthened by the observation that these mammalian territories of ~1,000 kbp [12], the supercoiled loops in interphasic *Drosophila* DNA of ~85 kbp [13], the yeast DNA rep-

licons [14] and transcription loops [1] all comprise about 10–15 genes on average. Moreover, imaging metazoan nuclei has unveiled thousands of small mRNA foci that each comprise 14–20 different transcription units [15, 16], which is fully consistent with 3-D clustering. A link between chromosome structure and transcriptional dynamics is also provided by the observation that mRNA synthesis inhibitors disperse DNA sequences originally confined to distinct regions and increase DNA loop mobility [17, 18]. Obviously, results with yeast must not be extrapolated to animal and plant cells without proper evidence. In the literature, 1-D contiguity, but not regular spacing, had been reported both for animal coexpressed genes and for yeast co-regulated genes [19, 20]. Table 1 summarizes the various cases reported in the literature [1, 2, 19, 20].

As far as prokaryotes are concerned, the 3-D clustering model predicts that transcription foci and long-range DNA loops should dynamically self-organize in the nucleoid, especially around the most active transcription units. Indeed, morphological approaches on live cells demonstrate discrete foci, each comprising hundreds of RNA polymerases engaged on the rRNA encoding operons in *Bacillus subtilis* [21] or containing TFs in *E. coli* [22]. Various structural and biochemical approaches suggest that, in actively growing *E. coli*

**Table 1.** Types of clustering between coregulated targets or between the gene encoding a TF and its targets

Organism	Distance: Clustering	Target/target		TF/target	
		1-D	3-D	1-D	3-D
Bacteria		+	+	+	+
Yeast		+	+	UA	UA
Animal		+	ND	UA	UA

+ = Observed; ND = not determined; UA = unapplicable.

cells, there would be ~50 independent loop domains per genome [23, 24], consistent with the transcriptomic results [2]. 3-D focusing is expected to depend on transcriptional activity. Indeed, starvation reduces the transcription rate and disperses the polymerases [21]. Along the same lines, in vivo structural data indicate that the number of loop domains drops as bacteria are grown on a poor medium [25] or enter the stationary phase [26].

### Explanatory Power

The concept of a genome-level transcriptional scheme provides a framework to explain heterogeneous and puzzling observations, including:

(1) The specificity of transcriptional and replicational regulations, despite the overwhelming abundance of unused protein binding motifs [14, 27–29]. There are even cases of TFs acting through low-affinity rather than high-affinity binding sites [30]. Indeed, in a solenoidal model of the genome, site position is a strong discriminant for effective site usage.

(2) The exquisite specificity of gene control by several TFs, despite its combinatorial intricacy. This regulatory interplay is indeed optimized by assigning the gene target to a cluster positioned for the correct set of TFs. Two corticoid receptors in rat neuron nuclei illustrate this possibility: these TFs were found together in only a subset of foci, and alone in others [31].

(3) The speed of coordinate transcriptional responses, given the size of the DNA molecule to be searched by TFs. The high synergy afforded by local concentration is indeed a powerful means to trigger ubiquitous responses with a minimal amount of additional molecules, that is within the shortest time [8, 9].

(4) The concerted reprogramming of gene expression upon an environmental change or over a cell cycle, given the high number of players (the case of transcriptional slowdown was already discussed). This versatility is built in the dynamic na-

ture of the foci, which maximizes responsiveness to the cell state [32].

(5) The high efficiency of DNA-related processes, despite extreme DNA compaction [33]. Compaction is indeed achieved in part by a mechanism that actually increases transcriptional efficiency. Moreover, the resulting DNA coaggregation favors intersegment transfer, thereby facilitating homologous recombination [34].

(6) The insensitivity of some bacterial promoters to their TFs when they are moved from their chromosomal loci to plasmids [35]. Moreover, some inversions of *E. coli* chromosomal regions are deleterious and induce an abnormal nucleoid structure [36, 37]. The solenoidal model dictates that site position must be respected for full functionality.

(7) The fact that yeast ‘heterochromatin barriers’ which block the spread of transcriptionally silent chromatin consist of multiple binding sites for various TFs [38, reviewed in 39]. Such multiple sites for any of those TFs would synergistically resist silencing thanks to clustering.

(8) The effect of TF dosage on both gene expression and chromatin structure or chromosome position, such as during the cell cycle [40], in response to interferon [41], in heterochromatic gene activation [42] or in position-effect variegation enhancement [e.g. 43]. An increased TF concentration would indeed recruit more DNA binding sites to its focus. Please note that this point applies only if the solenoidal concept turns out to extend to metazoans.

(9) The effect of transcription-related DNA motifs on gene position [44, 45]. Regulatory elements such as barriers, insulators, enhancers and enhancer blockers [44, 46] have been proposed to function by localized protein recruitment [47, reviewed in 48], for which one general strategy is suggested by chromosomal periodicity. Gene promoter can modulate alternative splicing [45], conceivably by moving towards the focus of its TFs, thereby loading the local set of splicing factors.

### Predictions of the Model

Testable consequences of the genome-level transcriptional scheme include:

(1) Similar regularities in other eukaryotes and prokaryotes, albeit with divergences reflecting differing transcription intensity, promoter organization or gene span. The testing will await the availability of sufficient data in other organisms, such as in nonconventional yeasts and in metazoans.

(2) A fine, periodic, positional effect on gene expression, particularly for genes with only one binding site for their TF, which therefore cannot increase local concentration by themselves. Genes at the optimal position for a given TF with respect to the solenoidal model would be regulated in a tighter fashion than genes that are farther from the optimal position, all other conditions being equal. This can be tested by systematic (or random, followed by assessment of position) displacement of a target gene and quantitative monitoring of its expression. In bacteria, this can additionally be tested by moving a TF-encoding gene and monitoring the expression of some of its targets.

(3) An alteration of the transcriptional status of nearby genes when the length of ‘junk’ DNA is locally altered in higher eukaryotes, if junk DNA acts as a spacer to improve transcriptional periodicity.

(4) A very precise path from DNA sequence to chromosomal morphogenesis. It predicts 1-D distant DNA regions that would cluster in 3-D. These predictions can be tested with tools such as chromosome conformation capture [49].

(5) The heterogeneous, multifocal distribution of a strong TF imaged either in active cells or following DNA spreading, and its redistribution upon environmental modification. Numerous small foci have already been observed for snRNA [50, 51] and mRNA [15, 16] genes. It is noteworthy that TFs were detected in snRNA foci. The mRNA foci comprise 14–20 different transcription units each. These foci are pre-

dicted to be elongating assemblies that were dynamically clustered at the time of transcription initiation because they share at least one common TF.

(6) Recognition of genuine protein binding sites among those detected through sequence analysis or through chromatin immunoprecipitation, guided by positional information. Predicted false positives, being in small number, can then be tested at the bench.

(7) A phylogenetic route for chromosome evolution through related species, based on transcription-induced periodicity. For instance, shortly after chromosomal splitting, both derivatives should retain an identical period; later, rearrangements should differentially alter their periods.

## Discussion

The present model, like Cook's recent model [52], points to mRNA transcription as a major architect of the functional nuclear structure. It is therefore interesting to note that these two models independently originated from very different experimental sources. Cook's model is primarily anchored in morphological studies, while this model relies on large-scale transcriptomic data. They both share a biochemical substratum as well. With respect to other DNA-related processes, Cook [52] proposed that transcription may also drive pairing between homologous chromosomes, while we showed that this transcriptional scheme constrains yeast replication origins [1], which in turn colocalize with long terminal repeats and other genomic features [14].

However, the two models have significant differences as well. A first difference is that only the present model, because it directly couples chromosomal structure to data bearing on DNA sequence, allows very accurate predictions, in particular on the respective spatial positioning of remote DNA segments. A second difference is that Cook emphasizes transcription elongation as the critical organizer, while the tran-

scriptomic data assign de facto the top role to transcription initiation. Although these views are mechanistically very different, they are not mutually inconsistent. Indeed, it may reasonably be assumed that once initiation has dynamically clustered coregulated targets, the latter will remain together in the same 'transcription factory' during elongation. A third difference is that these two models provide different physicochemical driving forces for the phenomenon (and consequential evolutionary rationale). We propose a dynamic cross-linking of bivalent TFs and polyvalent target genes, which have previously been shown to be effective at the intragenic level [8, 9]. Instead, Cook suggests that RNA-polymerase aggregation may provide the driving force, based on in vitro studies of the bacterial enzyme [53], but it is not clear how the results of such studies may be extrapolated to the in vivo context. A fourth difference is that Cook measures the chromatin loops in nucleotides (~85 kbp), while we would rather measure them in number of genes (~10–20 genes) when it comes to the interspecies comparison.

## Acknowledgments

This study was supported by funding from CNRS and Génopole®.

## References

- 1 Képès F: Periodic epi-organization of the yeast genome revealed by the distribution of promoter sites. *J Mol Biol* 2003;329:859–865.
- 2 Képès F: Periodic transcriptional organization of the *E. coli* genome. *J Mol Biol* 2004;340:957–964.
- 3 Shen-Orr SS, Milo R, Mangan S, Alon U: Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat Genet* 2002;31:64–68.
- 4 Guelzim N, Bottani S, Bourgine P, Képès F: Topological and causal structure of the yeast transcriptional regulatory network. *Nat Genet* 2002;31:60–63.
- 5 Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, Volkert TL, Wilson CJ, Bell SP, Young RA: Genome-wide location and function of DNA binding proteins. *Science* 2000;290:2306–2309.

- 6 Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, Zeitlinger J, Jennings EG, Murray HL, Gordon DB, Ren B, Wyrick JJ, Tagne JB, Volkert TL, Fraenkel E, Gifford DK, Young RA: Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 2002;298:799–804.
- 7 Miller OL Jr, Hamkalo BA, Thomas CA Jr: Visualisation of bacterial genes in action. *Science* 1970; 169: 392–395.
- 8 Müller-Hill B: The function of auxiliary operators. *Mol Microbiol* 1998;29:13–18.
- 9 Dröge P, Müller-Hill B: High local protein concentrations at promoters: Strategies in prokaryotic and eukaryotic cells. *Bioessays* 2001;23:179–183.
- 10 Jun S, Bechhoefer J: Role of polymer loops in DNA replication. *Phys Canada* 2003;59:85–92.
- 11 Cui Y, Bustamante C: Pulling a single chromatin fiber reveals the forces that maintains its higher-order structure. *Proc Natl Acad Sci USA* 2000;97:127–132.
- 12 Cremer T, Cremer C: Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* 2001;2:292–301.
- 13 Benyajati C, Worcel A: Isolation, characterization, and structure of the folded interphase genome of *Drosophila melanogaster*. *Cell* 1976;9:393–407.
- 14 Wyrick JJ, Aparicio JG, Chen T, Barnett JD, Jennings EG, Young RA, Bell SP, Aparicio OM: Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*: High-resolution mapping of replication origins. *Science* 2001;294:2357–2360.
- 15 Jackson DA, Iborra FJ, Manders EM, Cook PR: Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. *Mol Biol Cell* 1998;9:1523–1536.
- 16 Pombo A, Hollinshead M, Cook PR: Bridging the resolution gap: Imaging the same transcription factories in cryosections by light and electron microscopy. *J Histochem Cytochem* 1999;47:471–480.
- 17 Haaf T, Ward DC: Inhibition of RNA polymerase II transcription causes chromatin decondensation, loss of nucleolar structure, and dispersion of chromosomal domains. *Exp Cell Res* 1996;224:163–173.
- 18 Chubb JR, Boyle S, Perry P, Bickmore WA: Chromatin motion is constrained by association with nuclear compartments in human cells. *Curr Biol* 2002;12: 439–445.
- 19 Wagner A: Genes regulated cooperatively by one or more transcription factors and their identification in whole eukaryotic genomes. *Bioinformatics* 1999;15:776–784.
- 20 Roy PJ, Stuart JM, Lund J, Kim SK: Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans*. *Nature* 2002;418:975–979.
- 21 Lewis PJ, Thaker SD, Errington J: Compartmentalization of transcription and translation in *Bacillus subtilis*. *EMBO J* 2000;19:710–718.
- 22 Talukder AA, Hiraga S, Ishihama A: Two types of localization of the DNA-binding proteins within the *Escherichia coli* nucleoid. *Genes Cells* 2000;5:613–626.
- 23 Pettijohn DE: The Nucleoid; in Neidhardt FC (ed): *The Nucleoid*, ed 2. Washington, American Society of Microbiology, 1996, vol 1, pp 158–166.

- 24 Brunetti R, Prosseda G, Beghetto E, Colonna B, Micheli B: The looped domain organization of the nucleoid in histone-like protein defective *Escherichia coli* strains. *Biochimie* 2001;83:873–882.
- 25 Sinden RR, Pettijohn DE: Chromosomes in living *Escherichia coli* cells are segregated into domains of supercoiling. *Proc Natl Acad Sci USA* 1981;78:224–228.
- 26 Staczek P, Higgins NP: Gyrase and Topo IV modulate chromosome domain size in vivo. *Mol Microbiol* 1998;29:1435–1448.
- 27 Lieb JD, Liu X, Botstein D, Brown PO: Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. *Nat Genet* 2001;28:327–334.
- 28 Simon I, Barnett J, Hannett N, Harbison CT, Rinaldi NJ, Volkert TL, Wyrick JJ, Zeitlinger J, Gifford DK, Jaakkola TS, Young RA: Serial regulation of transcriptional regulators in the yeast cell cycle. *Cell* 2001;106:697–708.
- 29 Wingender E, Dietze P, Karas H, Knüppel R: TRANSFAC: A database on transcription factors and their DNA binding sites. *Nucleic Acids Res* 1996;24:238–241.
- 30 Wright WE, Funk WD: CASTing for multicomponent DNA-binding complexes. *Trends Biochem Sci* 1993;18:77–80.
- 31 van Steensel B, Van Binnendijk EP, Hornsby CD, Van der Voort HT, Krozowski ZS, De Kloet ER, Van Driel R: Partial colocalization of glucocorticoid and mineralocorticoid receptors in discrete compartments in nuclei of rat hippocampus neurons. *J Cell Sci* 1996;109:787–792.
- 32 Norris V: Modelling *Escherichia coli*. The concept of competitive coherence. *C R Acad Sci III* 1998;321:777–787.
- 33 Livolant F: Cholesteric organization of DNA in vivo and in vitro. *Eur J Cell Biol* 1984;33:300–311.
- 34 Berg OG, Winter RB, von Hippel PH: Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* 1981;20:6929–6948.
- 35 Martin RG, Rosner JL: Genomics of the marA/soxS/rob regulon of *Escherichia coli*: Identification of directly activated promoters by application of molecular genetics and informatics to microarray data. *Mol Microbiol* 2002;44:1611–1624.
- 36 Francois V, Louarn J, Patte J, Rebollo JE, Louarn J-M: Constraints in chromosomal inversions in *Escherichia coli* are not explained by replication pausing at inverted terminator-like sequences. *Mol Microbiol* 1990;4:537–542.
- 37 Guijo MI, Patte J, del Mar Campos M, Louarn J-M, Rebollo JE: Localized remodeling of the *Escherichia coli* chromosome: The patchwork of segments refractory and tolerant to inversion near the replication terminus. *Genetics* 2001;157:1413–1423.
- 38 Yu Q, Qiu R, Foland TB, Griesen D, Galloway CS, Chiu YH, Sandmeier J, Broach JR, Bi X: Rap1p and other transcriptional regulators can function in defining distinct domains of gene expression. *Nucleic Acids Res* 2003;31:1224–1233.
- 39 Bi X, Broach JR: Chromosomal boundaries in *S. cerevisiae*. *Curr Opin Genet Dev* 2001;11:199–204.
- 40 Tumber T, Belmont A: Interphase movements of a DNA chromosome region modulated by VP16 transcriptional activator. *Nat Cell Biol* 2001;3:134–139.
- 41 Volpi EV, Chevret E, Jones T, Vatcheva R, Williamson J, Beck S, Campbell RD, Goldsworthy M, Powis SH, Ragoussis J, Trowsdale J, Sheer D: Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. *J Cell Sci* 2000;113:1565–1576.
- 42 Lundgren M, Chow CM, Sabbattini P, Georgiou A, Minaee S, Dillon N: Transcription factor dosage affects changes in higher order chromatin structure associated with activation of a heterochromatic gene. *Cell* 2000;103:733–743.
- 43 Seum C, Spierer A, Pauli D, Szidonya J, Reuter G, Spierer P: Position-effect variegation in *Drosophila* depends on dose of the gene encoding the E2F transcriptional activator and cell cycle regulator. *Development* 1996;122:1949–1956.
- 44 Francastel C, Walters MC, Groudine M, Martin DI: A functional enhancer suppresses silencing of a transgene and prevents its localization close to centromeric heterochromatin. *Cell* 1999;99:259–269.
- 45 Cramer P, Caceres JE, Cazalla D, Kadener S, Muro AF, Baralle FE, Kornblihtt AR: Coupling of transcription with alternative splicing: RNA pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer. *Mol Cell* 1999;4:251–258.
- 46 Gerasimova TI, Byrd K, Corces VG: A chromatin insulator determines the nuclear localization of DNA. *Mol Cell* 2000;6:1025–1035.
- 47 Ptashne M, Gann A: Transcriptional activation by recruitment. *Nature* 1997;386:569–577.
- 48 Oki M, Kamakaka RT: Blockers and barriers to transcription: Competing activities? *Curr Opin Cell Biol* 2002;14:299–304.
- 49 Dekker J, Rippe K, Dekker M, Kleckner N: Capturing chromosome conformation. *Science* 2002;295:1306–1311.
- 50 Schul W, van Driel R, de Jong L: Coiled bodies and U2 snRNA genes adjacent to coiled bodies are enriched in factors required for snRNA transcription. *Mol Biol Cell* 1998;9:1025–1036.
- 51 Pombo A, Cuello P, Schul W, Yoon JB, Roeder RG, Cook PR, Murphy S: Regional and temporal specialization in the nucleus: A transcriptionally-active nuclear domain rich in PTE, Oct1 and PIKA antigens associates with specific chromosomes early in the cell cycle. *EMBO J* 1998;17:1768–1778.
- 52 Cook PR: Predicting three-dimensional genome structure from transcriptional activity. *Nat Genet* 2002;32:347–352.
- 53 Ishihama A: Subunit of assembly of *Escherichia coli* RNA polymerase. *Adv Biophys* 1981;14:1–35.