CellPress

Why the activity of a gene depends on its neighbors

Alexander Feuerborn and Peter R. Cook

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

Sixty years ago, the position of a gene on a chromosome was seen to be a major determinant of gene activity; however, position effects are rarely central to current discussions of gene expression. We describe a comprehensive and simplifying view of how position in 1D sequence and 3D nuclear space underlies expression. We suggest that apparently-different regulatory motifs including enhancers, silencers, insulators, barriers, and boundaries act similarly – they are active promoters that tether target genes close to, or distant from, appropriate transcription sites or 'factories'. We also suggest that any active transcription unit regulates the firing of its neighbors – and thus can be categorized as one or other type of motif; this is consistent with expression quantitative trait loci (eQTLs) being widely dispersed.

Position effects and gene regulation

In 1950 (three years before the description of the double helix), this is how Nobel laureate E.B. Lewis began a review [1]:

'That the effect of a gene may be dependent upon its position with respect to neighboring genes is now a well-established fact... This phenomenon of position effect... should throw light on the organization of the chromosomes as well as on the primary reactions of specific genes.' (Note: 'primary reactions' are now known as 'transcription').

In 2015, position effects (those due to position in 1D sequence space on a chromosome; see Glossary) are known to be commonplace; for example, expression levels of a reporter gene can vary $\sim 10^4$-fold when integrated at thousands of different sites around the mouse genome [2]. However, they are usually not central to current discussions of genome organization and gene regulation, where the underlying molecular mechanisms remain obscure. Nevertheless, all agree these mechanisms are complex, with $\sim 10^6$ sequences regulating only $\sim$23 000 human genes [3]. Moreover, regulators are diverse. We build on Lewis' 'fact' that 'position' is the key, and describe a comprehensive and simplifying view of how position (in 1D sequence and 3D nuclear space) determines gene expression (and vice versa). Our purpose is to specify more precisely what the underlying molecular mechanisms might be.

Some forces shaping structure

As pathologists know, genome architecture varies from cell to cell – even in clonal populations. At the molecular level, high-throughput chromosome conformation capture (Hi-C) applied to single mouse cells reveals that none share exactly the same genic contacts; however, some contacts are seen more often than others, and therefore the organization is non-random [4]. Time-lapse imaging of living human cells also shows the organization changes from moment to moment; a locus tagged with GFP might diffuse through a local volume (diameter 0.5–1 μm) for a minute or more (to contact briefly many other sequences), 'jump' to a neighboring volume the next (to contact others), and then become transiently immobilized [5,6].

Which proteins might stabilize specific contacts? We begin with transcription factors because they provide the necessary specificity. Many factors (either acting alone, or complexed with others) are 'bivalent' in the sense that they (or the complex) can bind to two different segments of DNA to form a loop. Box 1 illustrates three different ways they can stabilize loops, but only the first two require such bivalency. Any loops that are formed will persist for the order of seconds – the average residence time of a typical factor on DNA (again shown by GFP tagging [5]). However, engaged RNA polymerases can remain bound for longer (polymerase II takes ~10 min to transcribe a typical human gene of 30 kb), and this tight binding is specific in that it occurs throughout the transcription unit but not elsewhere. If two engaged polymerases are associated with other bivalent factors or complexes, then the same three ways can drive polymerases together. Since one-third of

Glossary

3C: chromosome conformation capture, a technique for assessing the proximity between two sequences on a chromosome in 3D nuclear space.
CTCF: CCCTC-binding factor, originally defined as a transcription factor.
eRNAs: transcripts encoded by enhancers.
eQTLs: expression quantitative trait loci.
DRB: 5,6-dichloro-1-β-D-ribo-furanosyl-benzimidazole, a transcriptional inhibitor.
GWAS: genome-wide association studies – the examination of many genetic variants in different individuals to see if any one variant is associated with a given phenotypic trait.
ENCODE: The Encyclopedia of DNA Elements.
Hi-C: a high-throughput variant of 3C.
HMR: hidden mating-type locus right, a locus controlling yeast mating type.
HUVECs: human umbilical vein endothelial cells.
NF-κB: nuclear factor κB, a transcription factor.
TNFα: tumor necrosis factor α, a cytokine.
Position effect: effects on expression of changing the location of a gene on a chromosome.
YY1: Yin Yang 1, usually considered a transcriptional repressor.

Corresponding author: Cook, P.R. (peter.cook@path.ox.ac.uk).
Keywords: boundary; domain; enhancer; expression quantitative trait loci; position effect; silencer.

0168-9525/
 http://dx.doi.org/10.1016/j.tig.2015.07.001

Box 1. Equilibria favoring looping

A typical transcription factor is present at ~1 nM, and many bind to others with equilibrium constants of $\sim 10^{-7}$ M; then, these numbers mean that <1% are instantaneously in protein:protein complexes (Figure IAi). However, if DNA with two cognate binding sites 10 kb apart is present, ~67% are in protein:protein complexes bound to DNA; DNA binding increases the local protein concentration, and thus interaction frequencies. Because such protein:protein complexes are 'bivalent', they can loop DNA [67] (Figure IAii).

Such clustering/looping is probably reinforced by a 'bridging-induced attraction' uncovered using (molecular dynamics) simulations of bivalent 'factors' binding to – and dissociating from – 'chromatin fibers' [68] (Figure IB). In the absence of explicit interactions between one factor and another, transiently bound factors nevertheless cluster spontaneously. Once a cluster happens to appear, it tends to persist; dissociating 'bridges' rebind nearby because the local concentration of binding sites is high (i.e., near other 'bridges').

Transcription factors are often found in larger complexes, and then 'depletion attraction' provides another force driving clustering [69] (Figure IC). In a crowded cell, small proteins (diameter <5 nm) continually bombard larger complexes (diameter 5–25 nm) from all sides. When two larger complexes come into contact, the small proteins are sterically prevented from entering the green volume between the two and thus cannot knock them apart; as a result, the small proteins exert a force equivalent to the osmotic pressure on opposite sides of the two larger complexes to keep them together (Figure ICi). If the larger complexes have DNA-binding sites, this 'attraction' can again stabilize loops (Figure ICii).

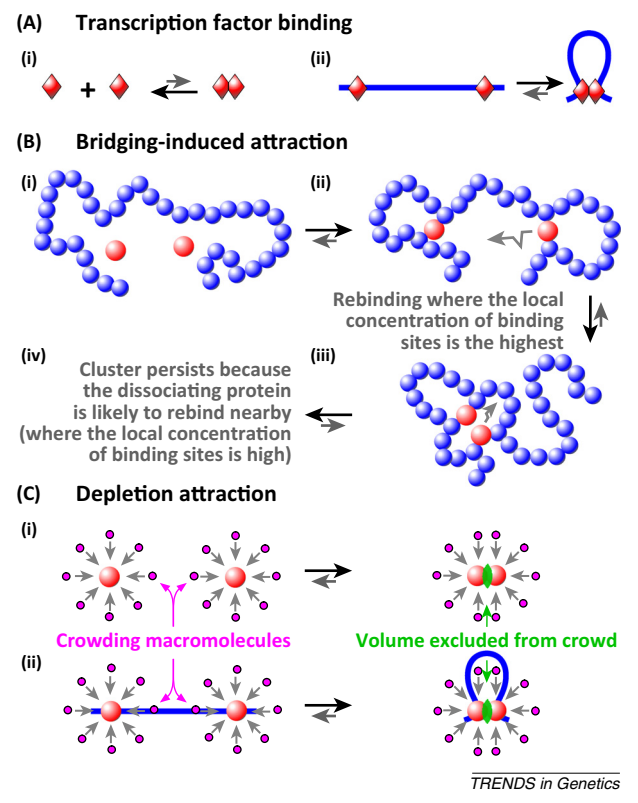

Figure I. Three equilibria favoring looping.

engaged polymerases are also 'paused'/'stalled' [7], these aggregates could persist for longer [7]. Hence, the system must either spend energy to prevent the clustering or – as seems likely – it goes with the flow and uses other familiar forces (charge interactions, H bonds, van der Waals, and hydrophobic forces) to organize the resulting structures.

Clusters of active polymerases – 'factories'

The forces described above fit comfortably with a model for genome organization in which a central architectural feature is a cluster of active polymerases – a 'transcription factory' – surrounded by loops [8,9] (Figure 1). We define such a factory as a site containing at least two polymerases (plus associated factors) active on at least two templates (to distinguish it from the case where two polymerases are active on one template). These factories contain high local concentrations that act through the law of mass action to drive production. For example, mammalian nuclei contain a 1 μM pool of polymerase II, but essentially all transcripts are made in factories where concentrations are ~1000-fold higher.

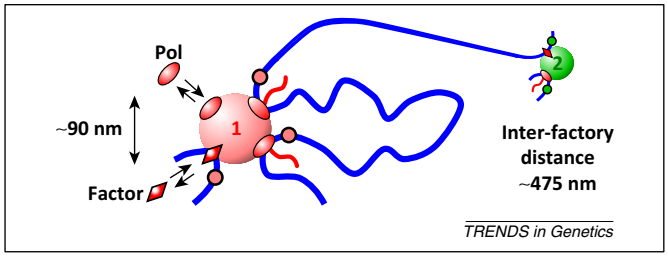

Figure 1. Transcription factories. Chromatin is tethered through clusters of polymerases/factors to two nucleoplasmic factories (1,2) which are rich in different factors. A typical factory is associated with ~16 loops (~eight tethered through active polymerases and ~eight through factors [9,22,23]). In this and subsequent figures, only a few attached loops are shown, polymorphic factories are represented as uniform spheres, and promoters (colored circles) tend to initiate in factories of the same color.

The first evidence for factories came when permeabilized human cells were incubated in bromouridine triphosphate (BrUTP) plus the other triphosphates required for transcription; after immuno-labeling, nascent BrRNA was seen in discrete sites [10]. These sites are so closely spaced they are difficult to resolve one from another by conventional microscopy, but clusters of polymerases [11] and appropriately tagged factors [12] have now been imaged in living cells using modern techniques; even so, the exact relationship of these clusters to active sites of transcription remains obscure. Factories have also been purified and their proteomes and transcriptomes analyzed – they contain the relevant polymerases, factors, RNA-binding proteins [13], and transcripts [14]. Significantly, the most frequently found contacts detected by chromosome conformation capture (3C)-based methods involve active transcription units [15–19] – the expected result if active units are tethered to factories.

Some car factories make Fords, others Hondas; do factories also specialize in transcribing specific gene sets? They do [9]. The nucleolus provides the prototypic example – a place where many rRNA genes are cotranscribed by polymerase I. Active polymerases II and III are also found in their own nucleoplasmic factories, and many different