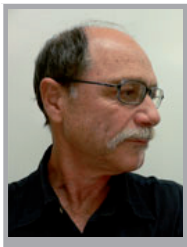
Nucleosome positioning, gene regulation and disease

"Genome-wide studies on DNA packaging and on how information is encoded in DNA to influence nucleosome arrangements may be beginning to reveal how enhancers function to regulate genes and how epigenetic modifications can bring about change."



Arnold Stein

Department of Biological Sciences, Purdue University, West Lafayette, IN 47906, USA
Tel.: +1 765 494 6546
Fax: +1 765 494 0876
steina@purdue.edu

Recent applications of genomic approaches to long-standing problems have set the stage for potential breakthroughs in gene regulation, and in providing new insights into disease mechanisms and possible intervention strategies. The details of how DNA is packaged in order to provide a compact organized state in the cell nucleus, while at the same time allowing both constitutive and tissue-specific gene regulation and other cellular functions, have eluded us. This difficulty is not surprising owing to the complexity of the problem. Nevertheless, many chromatin structure and nuclear organization models have been proposed over the years, and models tend to have a large influence on how we think about gene regulation and other cellular mechanisms. Several recent studies have serendipitously revealed some unexpected things about the organization of chromatin in the nucleus. Genome-wide studies on DNA packaging and on how information is encoded in DNA to influence nucleosome arrangements may be beginning to reveal how enhancers function to regulate genes and how epigenetic modifications can bring about change.

"...many chromatin structure and nuclear organization models have been proposed over the years, and models tend to have a large influence on how we think about gene regulation and other cellular mechanisms."

The following picture has emerged over the years. The nucleus appears to be well organized and chromosomes have territories [1]. There may be many chromatin fiber loops present [2]. The bulk of the chromatin (the protein–DNA complex in the nucleus) has generally been thought of as having a well-defined and uniform structure. For example, the solenoid model [3] has dominated the literature for more than 30 years. Recently, the somewhat modified 'interdigitated' solenoid model [4] or the very different 'zigzag' model [5], which resembles two side-by-side stacks of coins twisted about each other, have become popular [6]. Chromatin, whatever its structure is, consists of very long arrays of nucleosomes, the fundamental building blocks of chromatin. Each nucleosome contains 147 bp of DNA tightly wrapped 1.7-times around a protein spool made of two of each of the four core histones [7]. Each nucleosome is separated from its neighbor by a stretch of linker DNA, which can vary in length. In mammalian cells, linker DNA lengths vary from 18 to 80 bp [8]. In yeast (*Saccharomyces cerevisiae*), which has much shorter linkers than mammals, nearly 80% of the nucleosomes are in some way positioned with respect to the DNA sequence [9]. The extent of genome-wide nucleosome positioning in mammalian cells has not yet been determined.

The mechanism through which enhancer elements regulate genes from great distances undoubtedly involves chromatin. A popular idea is that the DNA between an enhancer and a transcriptional promoter loops out to allow interactions to occur between DNA-binding proteins that are far apart along the DNA [10]. If the initially inactive gene were packaged into a solenoid-like chromatin structure, activation would need to involve extensive remodeling so as to first unfold the condensed solenoid to make it flexible enough to loop. Although reasonable, this type of remodeling has not been clearly demonstrated. Alternatively, the nuclear address model postulates that there are active and inactive compartments in the nucleus [1]. Active genes are directed to active nuclear compartments by enhancer elements, which bind to proteins that are located in the active compartments. Active compartments of the nucleus may

future medicine part of fsg

be enriched in transcription factors and other proteins required for gene activation. It is fair to say that despite 30 years of effort it is not really known how enhancers act at a distance.

Also present in genomic DNA are insulator elements that serve to control or restrict enhancer activity [11]. An insulator placed between an enhancer and a promoter negates the activating effect of the enhancer, at least in artificial sequence contexts. Recently, by assaying extremely conserved human and mouse sequences for enhancer activity, it has been suggested that there are many functionally redundant short (~200 bp) DNA sequences in the mammalian genome that can serve as enhancers [12]. The number of such sequences was increased further when DNA structural considerations, rather than just sequence, were taken into account [13]. It is likely that the number of insulator sequences in the genome is also large [14], although experimental estimates using the methods applied to enhancers have not yet been made. In *Drosophila*, many 'boundary elements', which appear to limit the domain over which enhancers act, exist. Boundary elements also function as insulators [15]. Interestingly, the proteins that bind to boundary elements interact, apparently causing the DNA (really chromatin) between the elements to form loops [15].

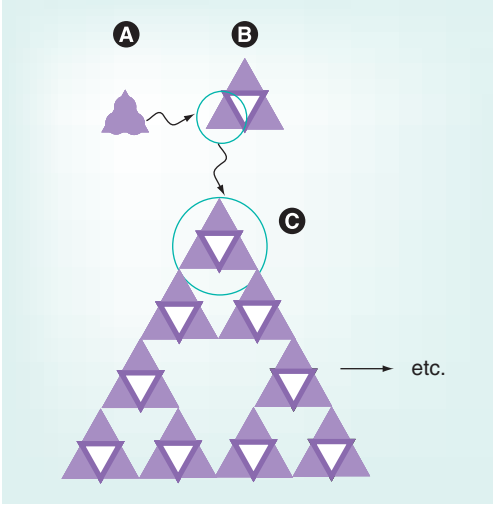


Figure 1. Fractal nucleosome packing arrangement using the Sierpinski triangle as an example. Only a small portion of the nucleus is shown. **(A)** Three nucleosomes form a small triangle. **(B)** A larger triangle containing nine nucleosomes is made from three small triangles. **(C)** A still larger triangle containing 27 nucleosomes is made from three triangles as shown in **(B)**; then another even larger triangle containing 81 nucleosomes is made from three of those triangles and so on. The encircled triangles correspond to the whole previous structure (wavy arrows). All of the triangles are self-similar, and are less dense in their central regions than in their peripheral regions. Packaging in this way can continue until the nucleus is filled.

> *"... there are many functionally redundant short (~200 bp) DNA sequences in the mammalian genome that can serve as enhancers."*

A method called chromosome conformation capture was developed to detect and quantitate the extent to which one region of chromatin is proximal to another region [10]. The chromatin in cells is fixed, the DNA is cut with a restriction enzyme, the chromatin is then treated with DNA ligase and the extent of ligation between any two nonadjacent regions is determined by PCR. Thus, DNA looping interactions between any two specific regions can be assessed. Recently, a modification of this technique using massively parallel sequencing was applied to detect all of the looping interactions in the genome (at a resolution of ~1 Mb) in a human lymphoblastoid cell line [16]. Somewhat serendipitously, an interesting pattern emerged revealing folding principles of the human genome. These experiments revealed the presence of two compartments in the nucleus, which were referred to as 'open' and 'closed' chromatin. The open chromatin possessed fewer looping interactions at any particular linear DNA distance than the closed chromatin. In addition, evidence of a fractal packaging pattern was obtained whereby a globular chromatin structure is successively folded into larger and larger globules that are self-similar on all scales. This packaging scheme was termed a fractal globule. The idea is illustrated in Figure 1 using the well-known Sierpinski triangle fractal. The actual fractal pattern and chromatin structures are not known, but the authors suggest that DNA folding may resemble a Peano curve (see Figure 2), which can fill an entire space without ever crossing itself [17]. Thus, the fractal globule is a knot-free arrangement that enables dense packing while allowing particular loci to easily fold, unfold or rearrange [16].

The fractal globule packaging of the genome, consisting of many DNA loops, taken together with the presence of many enhancer-like sequences and many insulators that interact with each other suggest that genes may be regulated as illustrated in Figure 2. Insulators,

which might more appropriately be called loop configuring elements, serve to stabilize and to rearrange chromatin loops by binding proteins that can interact with each other. To activate a gene, the loops are arranged so that an enhancer is proximal (in space) to the gene promoter. Alternatively, to insure the gene is inactive, loops are arranged so that enhancers are kept away from the promoter. Chromatin loop arrangements are induced by histone modifications, which can influence chromatin structure, [18,19], and by DNA methylation, which regulates protein binding to insulators [11]. The details of these processes are currently unknown.

> *"...the fractal globule is a knot-free arrangement that enables dense packing while allowing particular loci to easily fold, unfold or rearrange."*

The chromatin structures that give rise to the fractal globule packaging may be consequences of the nucleosome arrangements along the DNA. It has recently been shown that *in vitro* the chromatin higher order structure is strongly influenced by the nucleosome arrangement. Tandemly repeated nucleosome positioning sequences having different uniform internucleosome spacing, made using synthetic DNA, folded into dramatically different structures under identical conditions [20]. In cellular chromatin, DNA linkers are not uniform, and nonsolenoid models predict that many different intricately folded chromatin structures are possible [21,22].

Nucleosome arrangements are unlikely to be random. There is evidence that nucleosome positioning, resulting from DNA sequence-dependent bending around the histone core of the nucleosome, leads to different nucleosome arrangements [23] which correlate with chromosome function [24]. Although claims that there is a genomic code for nucleosome positioning [25] appear to be exaggerated [26], the extent of nucleosome positioning encoded in genomic DNA may be sufficient to direct the formation of chromatin structures that are consistent with the fractal globule chromatin packaging in the nucleus. It is also plausible that regions of chromatin do not have to be glued together by specific protein–protein interactions. Rather, self-similar structures might interact with each other through multiple weak interactions, such as van der Waals interactions. Chromatin remodeling would be expected to disrupt such interactions.

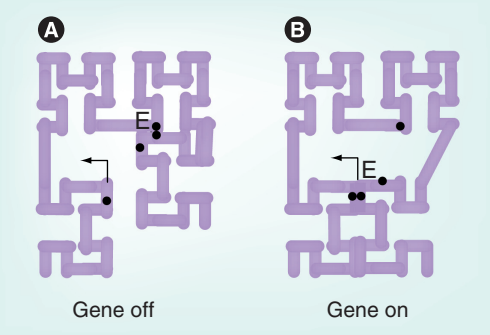


Figure 2. Remodeled chromatin Peano curves showing how 'insulators', enhancers and programmed chromatin structures cooperate to achieve gene regulation. In this hypothetical example, the chromatin fibers (thick lines) are programmed by the underlying DNA sequence to pack like a (slightly perturbed) Peano curve (Hilbert construction), a densely packed fractal pattern made by repeatedly duplicating staple-like shapes and then connecting them [17]. In other regions of the nucleus, the chromatin may be programmed to pack in a different way. The small filled circles represent loop configuring elements (insulators); E denotes the location of an enhancer element. **(A)** Gene off. The loop configuring elements interact through DNA-binding proteins to keep the enhancer away from the promoter region in front of the gene (shown by the right-angle arrow). **(B)** Gene on. Chromatin is remodeled, and the loop configuring elements then interact differently to place the enhancer close to the promoter.
E: Enhancer.

Many human syndromes are associated with defects in transcriptional regulation [27]. These include mutations in chromatin remodeling factors. It is also becoming increasingly clear that epigenetic changes are associated with disease states [28,29]. The new picture of DNA packaging provided by genome-wide experiments, together with the realization that information exists in genomic DNA that influences nucleosome positioning, nucleosome arrangements and chromatin structure, should provide new insights into the mechanisms involved in gene regulation and human disease.

Financial & competing interests disclosure

The author has no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Bibliography

1 Misteli T: Beyond the sequence: cellular organization of genome function. *Cell* 128, 787–800 (2007).

2 Hart CM, Laemmli UK: Facilitation of chromatin dynamics by SARs. *Curr. Opin. Genet. Dev.* 8, 519–525 (1998).

3 Finch JT, Klug A: Solenoidal model for superstructure in chromatin. *Proc. Natl Acad. Sci. USA* 73, 1897–1901 (1976).

4 Robinson PJJ, Fairall L, Huynh VAT, Rhodes D: EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure. *Proc. Natl. Acad Sci. USA* 103, 6506–6511 (2006).

5 Dorigo B, Schalch T, Kulangara A, Duda S, Schroeder RR, Richmond TJ: Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science* 306, 1571–1573 (2004).

6 Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P: *Molecular Biology of the Cell* (*5th Edition)*. Garland Science, NY, USA (2008).

7 Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ: Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–260 (1997).

8 Strauss F, Prunell A: Nucleosome spacing in rat liver chromatin. A study with exonuclease III. *Nucleic Acids Res.* 10, 2275–2293 (1982).

9 Lee W, Tillo D, Bray N *et al.*: A high-resolution atlas of nucleosome occupancy in yeast. *Nat. Genet.* 39, 1235–1244 (2007).

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

11 Bell AC, West AG, Felsenfeld G: Insulators and boundaries: versatile elements in the eukaryotic genome. *Science* 291, 447–450 (2001).

12 Visel A, Prabhakar S, Akiyama JA *et al.*: Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nat. Genet.* 40, 158–160 (2008).

13 Parker SCJ, Hansen L, Abaan HO, Tullius TD, Margulies EH: Local DNA topography correlates with functional noncoding regions of the human genome. *Science* 324, 389–392 (2009).

14 Wallace JA, Felsenfeld G: We gather together: insulators and genome organization. *Curr. Opin. Genet. Dev.* 17, 400–407 (2007).

15 Maeda RK, Karch F: The bithorax complex of *Drosophila*: an exceptional Hox cluster. *Curr. Top. Dev. Biol.* 88, 1–33 (2009).

16 Lieberman-Aiden E, van Berkum NL, Williams L *et al.*: Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–294 (2009).

17 Mandelbrot BB: *The Fractal Geometry Of Nature* (*2nd Edition)*. WH Freeman, NY, USA (1983).

18 Shogrer-Knaak M, Ishii H, Sun J, Pazin MJ, Davie JR, Peterson C: Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* 311, 844–847 (2006).

19 Robinson PJJ, An W, Routh A *et al.*: 30 nm chromatin fiber decompaction requires both H4-K16 acetylation and linker histone eviction. *J. Mol. Biol.* 381, 816–825 (2008).

20 Routh A, Sandin S, Rhodes D: Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. *Proc. Natl Acad. Sci. USA* 105, 8872–8877 (2008).

21 Woodcock CL, Grigoryev SA, Horowitz RA, Whitaker N: A chromatin folding model that incorporates linker variability generates fibers resembling the native structures. *Proc. Natl Acad. Sci. USA* 90, 9021–9025 (1993).

22 Engelhardt M: Choreography for nucleosomes: the conformational freedom of the nucleosomal filament and its limitations. *Nucleic Acids Res.* 35, E106 (2007).

23 Cioffi A, Fleury TJ, Stein A: Aspects of large-scale chromatin structures in mouse liver nuclei can be predicted from the DNA sequence. *Nucleic Acids Res.* 34, 1974–1981 (2006).

24 Takasuka T, Cioffi A, Stein A: Sequence information encoded in DNA that may influence long-range chromatin structure correlates with human chromosome functions. *PLoS One*, 3, E2643 (2008).

25 Segal E, Fondufe-Mittendorf Y, Chen L *et al.*: A genomic code for nucleosome positioning. *Nature* 442, 772–778 (2006).

26 Stein A, Takasuka TE, Collings CK: Are nucleosome positions *in vivo* primarily determined by histone-DNA sequence preferences? *Nucleic Acids Res.* 38, 709–719 (2010).

27 Bina M, Demmon S, Pares-Matos EI: Syndromes associated with *Homo sapiens* pol II regulatory genes. *Prog. Nucleic Acid Res. Mol. Biol.* 64, 171–219 (2000).

28 Cho KS, Elizondo LI, Boerkoel CF: Advances in chromatin remodeling and human disease. *Curr. Opin. Genet. Dev.* 14, 308–315 (2004).

29 Egger G, Liang G, Aparicio A, Jones PA: Epigenetics in human disease and prospects for epigenetic threrapy. *Nature* 429, 457–463 (2004).