Methods for Identifying Higher-Order Chromatin Structure

Samin A. Sajan and R. David Hawkins

Department of Medicine (Division of Human Genetics) and Department of Genome Sciences, University of Washington, Seattle, Washington 98195; email: rdhawk@uw.edu

Annu. Rev. Genomics Hum. Genet. 2012. 13:59–82

First published online as a Review in Advance on June 6, 2012

The *Annual Review of Genomics and Human Genetics* is online at genom.annualreviews.org

This article's doi:
10.1146/annurev-genom-090711-163818

1527-8204/12/0922-0059$20.00

Keywords

3C, 4C, 5C, Hi-C, ChIA-PET, FAIRE

Abstract

Eukaryotic genomic DNA is combined with histones, nonhistone proteins, and RNA to form chromatin, which is extensively packaged hierarchically to fit inside a cell's nucleus. The nucleosome—comprising a histone octamer with 147 base pairs of DNA wrapped around it—is the initial level and the repeating unit of chromatin packaging, which electron microscopy first made visible to the human eye as "beads on a string" nearly four decades ago. The mechanism and nature of chromatin packaging are still under intense research. Recently, classic methods like chromatin immunoprecipitation and digestion with deoxyribonuclease and micrococcal nuclease have been combined with high-throughput sequencing to provide detailed nucleosome occupancy maps, and chromosome conformation capture and its variants have revealed that higher-order chromatin structure involves long-range loop formation between distant genomic elements. This review discusses the methods for identifying higher-order chromatin structure and the information they have provided on this important topic.

INTRODUCTION

Nucleosome: the fundamental repeating unit of chromatin compaction, containing a histone octamer with 147 bp of DNA wrapped around it

Primary structure of chromatin: 11-nm chromatin fiber made of several nucleosomes in a row that under an electron microscope resemble beads on a string

Secondary structure of chromatin: 30-nm chromatin fiber containing a supercoiled 11-nm fiber with a zigzag arrangement of nucleosomes and linker DNA crisscrossing in between

Chromatin loops: higher-order chromatin structures formed by protein-mediated interactions between distal genomic regions such that the intervening DNA is looped out

Tertiary structure of chromatin: chromatin loops (and nucleosome plates) formed by the bending of the 30-nm fiber, a process aided by nonhistone proteins

Chromatin is composed of chromosomal DNA complexed with histones, nonhistone proteins (including structural and transcription factors), and RNA. Chromatin formation is a characteristic of eukaryotic genomes and enables the genome to be hierarchically packaged or condensed so that it can fit inside the nuclear space. Considering that naked DNA in a diploid human cell is approximately 2 m long, the extent of compaction at the level of the metaphase chromosome can be at least 10,000-fold. As if this were not a remarkable feat in itself, the nature of the compaction is such that it allows various factors to access certain regions of the genome to modulate processes such as gene transcription, DNA repair, recombination, and replication. And because these processes occur at different times and utilize different genes in different cell types in different environmental conditions, the chromatin structure is considered highly dynamic. A comprehensive knowledge of the structural features and dynamics of chromatin is therefore essential for understanding all cellular processes that involve DNA. Although this is quite a challenging task, it is nevertheless a worthwhile one to pursue. Indeed, much progress has been made in the past decade or so that has shed new light on this topic. In this review, we discuss methods that have enabled the discovery of higher-order chromatin structure as well as what we have learned from the knowledge gleaned through the application of these methods.

Chromosomes visible under a light microscope are obviously higher-order structures, with metaphase chromosomes representing the highest order that can be achieved physiologically. In the context of chromatin, the term higher-order structure presumes that there is a lower-order structure as well. Almost 40 years have passed since the first images of chromatin's fundamental repeating unit, the nucleosome, were obtained by electron microscopy and a model for the lowest/primary structure of chromatin was proposed (38, 45, 58). The model was based on the classical "beads-on-a-string" electron micrographs as well as data from biochemical and X-ray diffraction studies (21, 45). The nucleosome consists of an octamer of four different histone proteins (two each of H2A, H2B, H3, and H4), around which approximately 147 base pairs (bp) of DNA are wrapped in approximately 1.7 superhelical turns (21, 49, 66).

Several nucleosomes in a row form what is often referred to as a beads-on-a-string fiber, simply because of its appearance under an electron microscope (58, 59). This fiber, 11 nm in diameter, is typically considered the primary level of chromatin organization (**Figure 1*a***). When histones H1 or H5, referred to as linker histones, are added to the 11-nm fiber, a more condensed fiber with a diameter of 30 nm emerges (67). This fiber is said to represent the secondary level of chromatin organization (**Figure 1*c***). The linker histones cover anywhere from 20 to 90 bp of DNA (called linker DNA) depending on the species and tissues in question, with humans being around the 40-bp mark (1). The 30-nm fiber is arranged such that the linker DNA zigzags between two stacks of such fibers (69). In this particular arrangement, which is dependent on the ionic strength of the environment, odd-numbered nucleosomes make contact with other odd-numbered nucleosomes and even-numbered nucleosomes make contact with other even-numbered nucleosomes (**Figure 1*c***).

Concrete knowledge regarding structures beyond the 30-nm fiber is currently lacking, but there is evidence showing that these fibers are arranged in loops that constitute the tertiary structure of chromatin (**Figure 1*d***). Metaphase chromosomes and other regions of interphase chromosomes are manifestations of such chromatin loops. When chromosomes are depleted of histone proteins, a halo consisting of many loops of DNA (30–90 kb long) anchored along the length of a core/scaffold can be seen under an electron microscope (19, 64) (**Figure 1*d***). Recent evidence from cryo-electron microscopy also appears to support the existence of interdigitating layers of irregularly organized nucleosomes

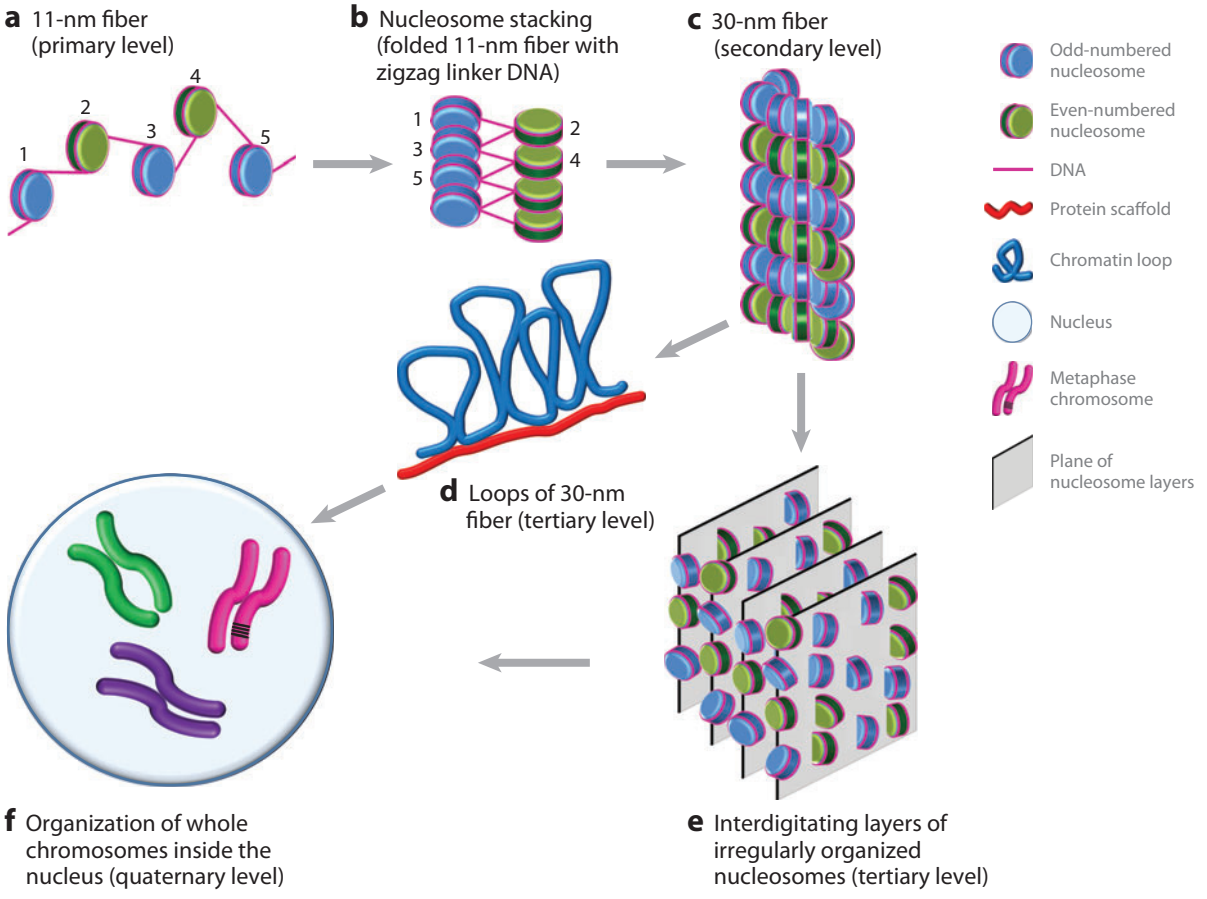


Figure 1

Different levels of chromatin compaction. (*a*) Multiple nucleosomes in a row form the 11-nm fiber that is the primary level of chromatin compaction. Alternating nucleosomes are depicted with blue and green surfaces. (*b*) The 11-nm fiber folds on itself to form two stacks/columns of nucleosomes such that odd-numbered nucleosomes interact with other odd-numbered nucleosomes and even-numbered nucleosomes interact with other even-numbered nucleosomes. The linker DNA zigzags between the two nucleosome stacks. (*c*) The folded 11-nm fiber forms a two-start helix to produce the 30-nm chromatin fiber that is the secondary level of compaction. (*d*) The 30-nm fiber twists further and forms a more compact fiber that is arranged in loops (*blue*), with some portions attached to a protein scaffold (*red*). This is one of the tertiary levels of compaction. (*e*) The 30-nm fiber may also result in the formation of interdigitating layers of irregularly oriented nucleosomes, particularly in metaphase chromosomes. Note that these plates do contain nucleosome fibers, but it is unclear whether they are 30-nm fibers or another type. Regardless, this is another tertiary level of compaction. (*f*) The quaternary level refers to the three-dimensional organization of entire chromosomes inside the nucleus and their relationships with one another as well as with the inner nuclear membrane. The black lines on the pink chromosome represent planes of nucleosome layers as viewed from above.

in metaphase chromosomes (8, 9, 26) (**Figure 1*e***). These, too, are considered to represent the tertiary level of chromatin packaging.

The quaternary structure of chromatin refers to the actual positioning of the chromosomes with respect to one another in the nucleus and with respect to the lamina of the inner nuclear membrane (**Figure 1*f***). It is known that expression of a gene is affected by its three-dimensional (3D) position within the nucleus, with the general consensus being that transcriptionally active genomic regions are further away from the nuclear periphery than those that are silent (80). The former

Quaternary structure of chromatin: the 3D positioning of chromatin domains relative to one another and to the nuclear lamina inside the nucleus

HETEROCHROMATIN

Heterochromatin refers to regions of the genome that are near the nuclear lamina, highly condensed, and transcriptionally silent. It can be facultative, meaning that transcriptional activity is spatially and temporally regulated through processes such as histone modification and DNA methylation, or constitutive, meaning that it is composed of repetitive elements occurring in blocks near centromeres and telomeres. Constitutive heterochromatin is responsible for position-effect variegation, a phenomenon whereby the expression of a gene located just outside its boundary is downregulated.

regions are called euchromatin and the latter heterochromatin; note that these terms refer more to levels of transcription and compaction than to levels of chromatin organization (see also sidebar Heterochromatin). Repression near the nuclear periphery appears to be mediated by the interaction of chromatin with lamin proteins of the inner nuclear membrane.

A consequence of loop formation is that genomic regions that are far apart on the linear DNA molecule are brought in close proximity to one another. This in turn can have profound effects on gene transcription because gene-distal enhancers can now directly interact with gene-proximal promoters. Enhancers can be thousands of kilobases away from their target genes in any direction (or even on a separate chromosome). Such enhancer-promoter interactions are brought about by sequence-specific factors that bind to DNA. A recently developed method called 3C (chromosome conformation capture; see below) and several other related techniques that employ genomics tools have enabled the identification of these interactions, thereby augmenting our knowledge of higher-order chromatin organization.

Epigenetic modifications also impact higher-order chromatin structure by altering the chemical properties of histones and certain DNA bases. For instance, acetylation of histones H3 and H4 on lysine residues makes them more negatively charged, thereby disrupting their electrostatic interactions with DNA and facilitating a less compact chromatin structure (73). Indeed, acetylation of histones is generally associated with actively transcribed regions and open chromatin regions. Methylation of histones and DNA does not affect the charge, but it does affect the binding of specific transcription factors that influence whether the genomic region will "open up" for transcription or not, thereby determining its compaction level.

Euchromatin: chromatin that is located away from the nuclear lamina, is generally less densely packed, and contains actively transcribed genes

Heterochromatin: chromatin that is near the nuclear lamina, tightly condensed, and transcriptionally silent

3C (chromosome conformation capture): intramolecular ligation-dependent method for identifying long-range chromatin interactions; several 3C variants exist, including 4C, 5C, 6C, Hi-C, and ChIA-PET

Sonication: breaking of DNA in chromatin using sound energy to obtain fragments between 140 and 1,000 bp suitable for most subsequent experiments

GENOME-WIDE METHODS FOR IDENTIFYING NUCLEOSOME POSITION AND NUCLEOSOME-FREE REGIONS

Are nucleosomes distributed randomly across the genome or are they placed in defined locations? Many studies have indicated that nucleosome organization is somewhat nonrandom and that some regions of the genome are more likely to contain nucleosomes than others. This nonrandom nucleosome positioning has clear implications for gene regulation and chromatin structure. Many experimental methods used to dissect the problem of nucleosome distribution across the genome involve a combination of steps, including the breaking up of chromatin through either sonication or digestion with nucleases [such as deoxyribonuclease I (DNaseI) and micrococcal nuclease (MNase)], chromatin immunoprecipitation (ChIP), and high-throughput sequencing.

Micrococcal Nuclease Digestion: MNase-Seq

MNase is the endonuclease of choice when isolating nucleosomes from chromatin after digesting away the linker DNA (57). Once a double-strand break is introduced, the nuclease digests the exposed ends until it runs into an obstruction such as a nucleosome or other proteins bound to DNA. Under limiting enzyme concentrations or time, the isolated DNA exhibits a "laddering" effect when resolved by gel electrophoresis, with each rung of the ladder corresponding to

the nucleosome-protected DNA. The smallest rung is usually approximately 150 bp (corresponding to mononucleosomal DNA), followed by 300 bp (dinucleosomal DNA), 450 bp (trinucleosomal DNA), and so on (**Figure 2*a***). The mononucleosomal DNA is then extracted from the agarose gel and processed accordingly either for hybridization to high-resolution tiling oligonucleotide microarrays or, as is more routine now, for high-throughput sequencing—i.e., MNase-seq (**Figure 2*a***). The use of paired-end sequencing can yield a short sequence from each end of the DNA wrapped around the nucleosome, providing a close approximation of the nucleosome-entering and nucleosome-exiting base positions.

DNA Sequence Context and Nucleosome Positioning

Studies of genome-wide nucleosome distribution have demonstrated that the underlying DNA sequence plays a significant role because the mechanical properties exhibited by some DNA sequences favor nucleosome formation. It turns out that sequences containing AA/TT/AT and GC dinucleotides at 10-bp intervals are most favorable for nucleosome formation (25, 39, 40, 48, 71). It is thought that this sequence pattern enables the DNA to acquire the amount of curvature necessary to accommodate nucleosomes. One study demonstrated the importance of DNA sequence for the formation of nucleosomes by assembling nucleosomes on purified yeast genomic DNA in vitro (41). The study showed that the in vitro nucleosome map (which was dependent only on DNA) was very similar to in vivo maps generated under three different growth conditions. Using their in vitro data, the authors were able to compile a computational program whose predictions of nucleosome positions correlated well with in vivo genome-wide nucleosome maps of *Caenorhabditis elegans* (41, 81).

DNA methylation also influences nucleosome positioning (10). Analysis of nucleosomal DNA together with whole-genome DNA methylation data from *Arabidopsis* indicated that nucleosomal DNA was highly methylated relative to non-nucleosomal DNA (10, 11). The same trend was also found in humans. Also, both DNA methylation and nucleosomes were found to be highly enriched on exons as well as intron/exon boundaries. Furthermore, methylated DNA can be bound by methyl-CpG domain binding proteins that recruit additional factors that modify histones or remodel nucleosomes, thereby affecting chromatin structure (15).

Nucleosome Positioning over Promoters

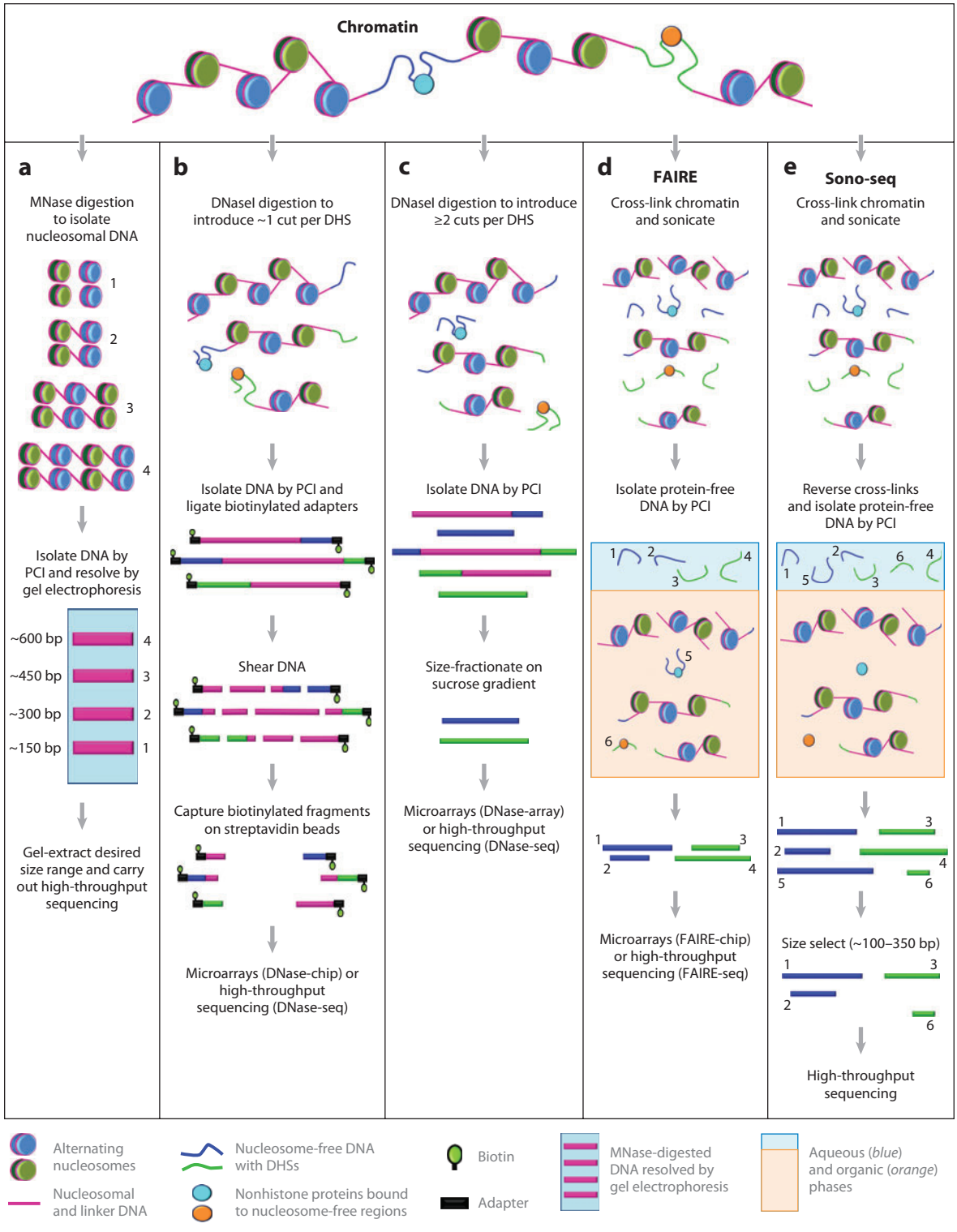
Studies in yeast and even human cells show that some gene promoters can be classified into two categories: open and occupied. Open promoters tend to lack a classical TATA box and have approximately 150 bp upstream of the transcription start site (TSS) that lack nucleosomes (51, 70, 89). This nucleosome-free region, at least in yeast, is flanked by two well positioned nucleosomes—the "−1" nucleosome upstream of the TSS and the "+1" nucleosome whose edge is just over the TSS. Interestingly, in open promoters of the fruit fly, the +1 nucleosome is approximately 50 bp more downstream of the TSS than it is in yeast (52). This likely reflects differences in how transcription is regulated in metazoans. The +1 nucleosome in yeast is almost directly at the TSS, which may influence transcription initiation by RNA polymerase II; in the fruit fly, this nucleosome is shifted more downstream, so that it interacts with the paused polymerase. Nucleosomes that are downstream of the +1 nucleosome and those that are upstream of the −1 nucleosome have a less well-defined positioning.

Open promoters generally regulate the expression of constitutive genes that do not require strict regulation, thereby allowing access to transcription factors. Occupied promoters, in contrast, are completely (or almost completely) occupied by nucleosomes. Thus, the TSS and any transcription factor binding sites within them are inaccessible. There may,

Deoxyribonuclease I (DNaseI): nuclease of choice for isolating nucleosome-depleted genomic DNA

Micrococcal nuclease (MNase): nuclease of choice for isolating nucleosome-occupied genomic DNA

Chromatin immunoprecipitation (ChIP): antibody-based enrichment of DNA fragments bound by a protein of interest, which can be identified by microarrays (ChIP-on-chip) or high-throughput sequencing (ChIP-seq)



however, be an entry site for so-called pioneer factors in the linker DNA or adjacent to the nucleosome; this site may even be buried within the nucleosome core and become accessible only after the action of certain chromatin remodelers. These promoters generally contain TATA boxes and are associated with tightly regulated genes such as those involved in the stress response and the cell cycle (3).

Nucleosome Positioning over Insulators

Insulators are DNA sequences whose function is to separate certain genomic regions from each other (enhancers from promoters, or euchromatin from heterochromatin). Integrative analyses of many genome-wide nucleosome mapping data sets showed that CTCF, the vertebrate insulator binding protein, binds to and protects approximately 60 bp of linker DNA and is flanked by 10 very well-positioned nucleosomes on either side (14, 22). These nucleosomes are highly enriched in the histone variant H2A.Z and in up to 11 different histone modifications (22). Surprisingly, the positions of these nucleosomes were better defined than those of nucleosomes near the TSS discussed above, further supporting the role of CTCF in influencing higher-order chromatin structure.

Nucleosome Positioning over Transcription Termination Sites

The 3′ ends of most genes are sites of RNA polymerase disassembly and transcript cleavage. There is normally a well-positioned nucleosome at these transcription termination sites as well as a nucleosome-free region beginning at the cleavage and polyadenylation signals (51, 72). The median length of these nucleosome-free regions at 3′ ends of genes is 173 bp in over 95% of examined genes (51). This region is not only where the transcription machinery disassembles but also where the antisense transcription initiation machinery assembles, thereby making these regions act as promoters for antisense transcription. Interestingly, at least in yeast, the terminating RNA polymerase II can "recycle" back to the promoter of the

Figure 2

Different methods used to determine regions of the genome that are either occupied or free of nucleosomes and/or other proteins. Chromatin containing nucleosomes (*blue* and *green*) and two nucleosome-free regions is shown at the top. Each nucleosome-free region, which also contains several DNaseI hypersensitive sites (DHSs), is shown as occupied by a nonhistone protein. (*a*) Micrococcal nuclease (MNase) digestion is used to determine regions of the genome that are covered by nucleosomes. MNase digestion will protect DNA covered by nucleosomes. Depending on the extent of digestion, it is possible to recover mono-, di-, tri-, or tetranucleosomes (or even higher multiples), shown by numbers 1, 2, 3, and 4, respectively. DNA is then isolated by phenol-chloroform-isoamyl alcohol (PCI) and resolved by gel electrophoresis. From bottom to top, each magenta line in the gel represents DNA protected by 1, 2, 3, or 4 nucleosomes, respectively, with a length of approximately 150 base pairs (bp) per nucleosome. Appropriately sized fragments are gel-extracted and sequenced to determine their identity and produce nucleosome occupancy maps. (*b*) DNaseI digestion is used to determine nucleosome-free regions. In this case, the digestion is carried out so as to introduce at most one cut per DHS, which is then followed by isolation of the DNA by PCI and ligation of biotinylated adapters. The DNA is then sheared, and biotinylated fragments are captured by streptavidin beads and either sequenced or hybridized to microarrays to determine their identity. (*c*) DNaseI digestion is carried out to introduce two or more cuts per DHS so as to release smaller fragments of the DHS. The DNA is isolated by PCI and size-fractionated on a sucrose gradient to isolate appropriately sized fragments, which are then either sequenced or hybridized to microarrays to determine their identity. (*d*) FAIRE (formaldehyde-assisted isolation of regulatory elements) is a method used to isolate protein-free regions of the genome. Here, chromatin is first cross-linked and sonicated. The DNA is then isolated by PCI to recover protein-free regions in the aqueous phase (*blue region*) and sequenced. The organic phase (*orange region*) retains all fragments bound by proteins. Fragments of the nucleosome-free regions are numbered. (*e*) Sono-seq is another method for isolating protein-free regions. It is similar to FAIRE but involves reversing the cross-links before cleaning the DNA with PCI so that proteins, particularly non-nucleosome proteins, that are loosely bound to the DNA will easily "fall off" and the fragments to which they were bound will partition in the aqueous phase and be recovered. All recovered fragments from 100 to 350 bp are then sequenced. Thus, although some common fragments may be captured by both FAIRE and Sono-seq (fragments 1, 2, and 3 in this figure), there are also some that are unique to each method (fragment 4 is unique to FAIRE and fragment 6 is unique to Sono-seq).

same gene via a looping mechanism whose physiological relevance is still unclear (51, 75).

DNaseI hypersensitive site (DHS): genomic region that is easily accessible and digested by DNaseI owing to a paucity of nucleosomes and nonhistone proteins

DNase-array, DNase-chip, and DNase-seq: methods to identify DHSs on a genome-wide scale using dense oligonucleotide microarrays (DNase-array and DNase-chip) or high-throughput sequencing (DNase-seq)

MNase-Seq "Version 2.0"

MNase digestion is preferably carried out without a formaldehyde cross-linking step to prevent a strong association of nonhistone proteins to DNA, thereby ensuring that only nucleosome-bound DNA will be protected. Nevertheless, a recent study has shown that it is possible to detect smaller DNA sequences (less than 100 bp) occupied by nonhistone proteins even without cross-linking by omitting a size-exclusion step found in most standard protocols (37). Isolating and sequencing DNA fragments of less than 100 bp enabled the identification of short, protected fragments enriched in known transcription factor motifs. Moreover, these sites largely overlapped nucleosome-free regions at regulatory elements. Thus, inclusion of short fragments from MNase digestion in sequencing libraries will provide genome-wide occupancy information pertaining to both nucleosomes and transcription factors.

DNaseI Digestion: DNase-Seq

In contrast to MNase digestion, DNaseI digestion identifies nucleosome-free DNA. It has been almost four decades since the first observation that transcriptionally active genes are preferentially digested by DNaseI (84). Shortly after this observation, regions of the genome that were even more sensitive to DNaseI digestion were discovered and referred to as DNaseI hypersensitive sites (DHSs). These were depleted in nucleosomes and were region specific, indicating an alteration of higher-order chromatin structure at these locations (86, 87). For the vast majority of the time since this discovery, researchers had to carry out labor-intensive techniques to detect DHSs on a locus-by-locus basis. Specifically, chromatin had to be digested with a limiting amount of DNaseI so that on average there was one cut per DHS of interest, the DNA had to be purified and digested with a restriction enzyme that had recognition sites flanking the DHS of interest, and finally the size distribution of the digested fragments had to be determined by Southern blotting using a radiolabeled probe, which in turn required a prior knowledge of the DNA sequence being examined (42, 63).

With the advent of genome tiling arrays, it became possible to interrogate DHSs on a genome-wide scale. The first such studies used techniques like DNase-chip and DNase-array (12, 68). DNase-chip involves digesting chromatin with DNaseI, ligating biotinylated linkers to DNA ends, fragmenting the DNA, capturing biotinylated fragments using streptavidin, labeling them with fluorescent dye, and hybridizing them on high-density oligonucleotide microarrays (12) (**Figure 2*b***). DNase-array relies on two cuts made by DNaseI to release a portion of the DHS, followed by size-selection of the fragments to isolate smaller fragments representing the DHS, labeling of these fragments with fluorescent dye, and hybridization on high-density oligonucleotide microarrays (68) (**Figure 2*c***). The two studies that first employed these methods both examined the occurrence of DHSs in the same 30 million bp of the human genome selected by the Encyclopedia of DNA Elements (ENCODE) consortium. More than 2,500 DHSs were identified; surprisingly, the majority of these were over 10 kb away from annotated genes, and their overall distribution was such that they formed large DHS superclusters separated by 100–400-kb regions of few or no DHSs (68). This likely reflects a high-order organization of chromatin structure representing active and inactive domains.

These two studies used a total of two human cell lines; shortly thereafter, another study examined additional cell lines (88). Almost 4,000 DHSs were found across all six cell lines used, and 22% of these were present in all cell lines (ubiquitous DHSs), mostly within 2 kb of a TSS. Of the ubiquitous DHSs that were distal from a TSS, the majority were bound by CTCF, suggesting a novel role for ubiquitous DHSs (88). Cell type–specific DHSs, in contrast, mostly overlapped enhancers and

correlated with specific gene expression and histone modifications. Importantly, almost 8% of the genome overlapped with the ~4,000 DHSs discovered, indicating that a large proportion of the genome is functional.

Whereas the first studies that employed genomics to identify DHSs used microarrays, the more common methods today use high-throughput sequencing (DNase-seq). Massively parallel signature sequencing was the first high-throughput sequencing method used to identify genome-wide DHSs in undifferentiated CD4+ cells (7, 13). Although this was a significant improvement over the use of microarrays representing only 1% of the ENCODE-selected human genome, and considering that ~5,000 DHSs were identified, the method was still labor intensive and involved cloning fragments into vectors (7, 13). Furthermore, it was low-throughput (230,000 sequenced tags total) compared with current Illumina sequencing, which yields tens and even hundreds of millions of tags. The first truly high-throughput sequencing-based identification of DHSs used a combination of DNase-chip and DNase-seq by employing Illumina and Roche/454 sequencing to generate a total of 18 million tags (6). The study identified ~95,000 DHSs in human primary CD4+ T cells, of which a surprisingly small fraction (~20%) were near the TSSs of known genes (6).

More recently, DNase-seq was used to characterize DHSs in five embryonic stages of fruit fly development. The study found that, in terms of the number and distribution of DHSs, the global chromatin architecture was similar in both early and late embryonic stages (77). However, this similarity was due to a dynamic chromatin landscape where changes were occurring at the level of individual DHSs, so that the overall genomic view appeared deceptively similar. One surprising finding was the presence of developmentally regulated localized access (or weak DHSs) along the bodies of protein-coding genes. Many of these genes, it turned out, were maternally contributed during early development prior to the onset of zygotic transcription (77). Thus, the presence of partially open chromatin on exons of these genes could potentially facilitate their timely activation in the zygote once the maternal contribution has been depleted or is no longer needed. This opens up a novel avenue of research in terms of transcriptional regulatory mechanisms applicable to a subset of genes during development.

FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements)

FAIRE (formaldehyde-assisted isolation of regulatory elements): method for isolating protein-free genomic regions by PCI and identifying them using oligonucleotide microarrays (FAIRE-chip) or high-throughput sequencing (FAIRE-seq)

FAIRE was developed in an attempt to identify all genomic regions that were nucleosome-depleted using few reagents and a minimum number of steps. Briefly, cells are first cross-linked using formaldehyde, the chromatin is isolated and fragmented by sonication, and DNA is isolated using phenol-chloroform-isoamyl alcohol (PCI) extraction. The idea is that chromatin fragments that are depleted in proteins—and hence representing active regions—will preferentially separate into the aqueous phase. Once isolated, these fragments can be analyzed either by hybridizing them to high-density oligonucleotide microarrays (FAIRE-chip), as was originally done, or by high-throughput sequencing (FAIRE-seq), as is more routine now (28, 55, 56, 82) (**Figure 2*d***). FAIRE-chip was first demonstrated to be an effective method for identifying active chromatin regions through use of a human foreskin fibroblast cell line and a microarray representing the ENCODE-selected 30 million bp of the genome (28). More than 1,000 FAIRE peaks were found, and they overlapped well with various marks of open chromatin such as DNaseI hypersensitivity, TSSs, and active promoter regions. Interestingly, 43% of the peaks were termed "orphans" as they failed to overlap with any of the annotations that were selected for comparison, leading the authors to speculate that these peaks were likely due to differences in cell lines used by different studies, lack of good annotation in these particular genomic regions, or novel functions that are not associated with DHSs or other previously described hallmarks of active chromatin (28).

Sono-seq: method for isolating protein-free genomic regions by sonicating chromatin and then size-selecting small fragments between 100 and 350 bp

Owing to its high-throughput nature, FAIRE-seq has enabled the identification of many more open chromatin regions in different cell types. One example is the finding of 80,000 such regions in human pancreatic islet cells, of which 3,300 were islet-specific regions associated with genes whose expression was also islet-specific (27). More important, the study discovered that one islet-specific open chromatin region contained the intronic single-nucleotide polymorphism (SNP) rs7903146 in the *TCF7L2* gene, which has been associated with type 2 diabetes (36). Application of FAIRE to islet cells from individuals heterozygous at this SNP showed that the T allele, which is the at-risk allele, contained more open chromatin than the C allele (27, 29). Indeed, the T allele exhibited higher luciferase reporter activity in islet-specific cell lines compared with the C allele, suggesting that the at-risk allele predisposes to disease by changing the local chromatin structure in islets.

Another example is a study that employed both DNase-seq and FAIRE to identify more than 870,000 open chromatin regions (representing ~9% of the genome) present in at least one out of seven different human cell lines (76). Interestingly, the authors observed that on average there was only one-third overlap between the top 100,000 DNaseI sites and the top 100,000 FAIRE sites from each cell line. Many of the DNaseI-only sites were near TSSs, whereas many of the FAIRE-only sites were in distal regions, particularly internal introns and exons. These differences may reflect the presence of specific regulatory proteins/complexes found in some regions such that they differentially affect the cutting and cross-linking abilities of DNaseI and formaldehyde, respectively. It may be challenging to capture some nucleosome-free regions by FAIRE if those regions happen to be bound tightly by other nonhistone proteins, which can become cross-linked to DNA by formaldehyde. Conversely, some nonhistone proteins may inhibit (sterically or enzymatically) the action of DNaseI so that these regions are not detected by DNase-seq. Hence, a combination of the two methods clearly provides a comprehensive view of open chromatin regions in the genome. This study also revealed that open regions common to all cell types tended to be near TSSs or bound by CTCF, whereas those that were unique to one cell type were far from TSSs and contained sequences that bind to factors that regulate cell identity (76). FAIRE has also been used to identify open chromatin regions in differentiating human adipocytes and in MCF-7 breast cancer cells (50, 82). The latter study showed that PBX1 is a pioneer factor that binds to certain regions of the genome, where it promotes a more open chromatin state. Some of these PBX1-mediated open chromatin regions then bind to estrogen receptor alpha, which acts as a transcription factor that mediates the progression of almost two-thirds of all breast cancers (50).

Sono-Seq

The Sono-seq method entails cross-linking chromatin with formaldehyde, sonicating the chromatin, reversing cross-links, isolating DNA by PCI, size-selecting fragments between 100 and 350 bp, and analyzing them by high-throughput sequencing (**Figure 2*e***). Thus, the only difference between a Sono-seq sample and the input control samples commonly used in ChIP experiments is that the former involves a size-selection step. The idea behind this method is that open chromatin regions tend to fragment more easily and readily relative to closed ones, and so the size-selection step is important because larger fragments (representing chromatin that is harder to fragment) are unlikely to be enriched with open regions. Limited sonication ensures that primarily nucleosome-free regions fragment first and are enriched in the size selection.

When this method was first demonstrated in HeLa cells, more than 21,000 strong Sono-seq peaks were found (2). Approximately two-thirds of these were less than 1,000 bp from peaks found by ChIP-seq using an RNA polymerase II antibody, indicating that they represented

promoter-proximal regions. The remaining one-third were located in CTCF binding sites, small RNA genes, and enhancers, demonstrating that Sono-seq is enriched for regions involved in a wide variety of biological functions (2). A comparison of Sono-seq peaks with DHSs generated by the study that first used DNase-chip showed that there was a significant overlap of Sono-seq peaks with promoter-proximal DHSs but not with distal DHSs (2, 12). A comparison with FAIRE peaks led to different results; in particular, there were no Sono-seq peaks overlapping FAIRE peaks, but there were many FAIRE peaks near the borders of Sono-seq peaks, indicating that Sono-seq and FAIRE are different from each other (2). One important methodological difference between Sono-seq and FAIRE is that in the former the cross-links are reversed before PCI extraction of DNA, whereas in the latter they are not (**Figure 2*d*,*e***). Thus, Sono-seq presumably captures some protein-bound fragments as well. However, in spite of its ability to capture many regulatory elements, Sono-seq has not been applied as widely as the other methods described above.

CHROMATIN IMMUNOPRECIPITATION: A POWERFUL METHOD FOR UNDERSTANDING CHROMATIN STRUCTURE THROUGH HISTONE MODIFICATIONS

The application of ChIP coupled with next-generation sequencing (ChIP-seq) to identify transcription factor binding sites and generate histone-modification maps has been the subject of numerous recent reviews and has proven to be an extremely important breakthrough in the field of chromatin biology (20, 62, 92). Therefore, we cover it here only in limited detail.

All of the above-mentioned methods benefited from the pioneering work of ChIP-chip, where labeled ChIP DNA is hybridized to microarrays. Both ChIP-chip and ChIP-seq have been instrumental in determining global maps of histone modifications that have an important relationship with nucleosome position, especially at regulatory elements. Nevertheless, whether the modifications are correlated with or causative of altered nucleosome positions remains to be determined.

Certain amino acids in the N-terminal tails of histones, and even some within their internal domains, are known to be posttranslationally modified. Several modifications have been reported, including methylation, acetylation, ubiquitination, and SUMOylation of lysine residues; methylation of arginine residues; isomerization of proline residues; ADP-ribosylation of glutamate residues; and phosphorylation of serine and threonine residues (the best-studied ones being acetylation and methylation of lysines) (5). These modifications affect the interactions of nucleosomes with one another as well as with other nonhistone DNA binding proteins, which in turn affect gene expression and ultimately the higher-order chromatin structure. Each of these modifications can be interrogated by ChIP-chip and ChIP-seq using antibodies that recognize the modified histone epitopes, allowing the generation of genome-wide histone-modification maps. H3K4me3 and H3K27me3 modifications, for example, are found in the promoters of actively transcribed and silent genes, respectively (31, 54, 61); the H3K36me3 mark, in contrast, is found primarily on exons of genes that are actively transcribed, hinting that splicing may also be tied to local chromatin structure (44). Acetylation of lysines is generally thought of as a mark of open chromatin at active regulatory elements and is found at promoters, enhancers, and actively transcribed genes as well (34, 83). Enhancers have been shown to contain H3K4me1, a mark that was then used to map 55,000 potential enhancers in multiple human cell lines (31, 32, 34, 35).

Although histone modifications are an important aspect of active and repressed chromatin structures, ChIP-chip and ChIP-seq maps are typically not informative about the underlying nucleosome position. However,

EXPERIMENTAL SUITABILITY OF THE METHODS FOR LONG-RANGE CHROMATIN INTERACTIONS

- 3C (chromosome conformation capture): Suitable for detecting a handful of long-range chromatin interactions. It requires a previous knowledge of interacting DNA sequences, as sequence-specific primers for amplification are needed.
- 4C (circular chromosome conformation capture and chromosome-conformation-capture-on-chip): Suitable for detecting long-range chromatin interactions that involve a particular genomic region of interest, commonly referred to as the bait.
- 5C (chromosome conformation capture carbon copy): Suitable for detecting a large number of long-range chromatin interactions. This is a high-throughput version of 3C, as it makes use of universal primers during amplification. However, a previous knowledge of interacting DNA sequences is still required to design a large number of specific primers to enable multiplex ligation-dependent probe amplification (MLPA).
- 6C (combined chromosome conformation capture ChIP cloning): Suitable for detecting a limited number of long-range chromatin interactions. It is similar to 3C but involves a ChIP step that reduces the complexity of the chromatin pool. It is also laborious, as it entails the cloning of fragments in bacteria followed by the examination of cloned DNA from individual bacterial colonies by restriction digestion and gel electrophoresis.
- Hi-C: Suitable for detecting a large number of long-range chromatin interactions globally. It does not require a previous knowledge of interacting DNA sequences.
- ChIA-PET (chromatin interaction analysis by paired-end tag sequencing): Suitable for detecting a large number of both long-range and short-range chromatin interactions globally. A ChIP step can also be included to reduce the complexity of the chromatin pool. It provides a quantitative measure of intermolecular ligation events that are false positives, making this a superior method.

recent advances in computational methods have capitalized on the high-resolution data generated from these experiments to identify nucleosome-free regions, especially at histone-modified regulatory elements (33, 90). A comparative analysis of data sets reveals nucleosome dynamics at these elements and identifies the nucleosome-free region within the corresponding histone-modification marked element.

METHODS FOR IDENTIFYING LONG-RANGE CHROMATIN INTERACTIONS

As mentioned above, genomic regions that are distal to one another on the linear strand of DNA can physically interact. This is achieved by the creation of chromatin loops—such as those between gene-distal enhancers and gene-proximal promoters—to regulate transcription. The locations and numbers of such loops, and hence the higher-order 3D chromatin structure, will undoubtedly be dynamic as they will depend on gene expression patterns within a particular cell at a particular time point.

The 3C method was the first to identify some of the long-range interactions that occur in yeast (16). Application of 3C showed that the interaction frequency of two loci is inversely proportional to the physical distance between them. The 3C data, when translated into 3D chromosomal distances using a polymer-based mathematical model, further revealed that the chromatin fiber is a flexible, freely jointed chain and that yeast chromosome III exists as a contorted ring with juxtaposed telomeres (16). Below, we describe the 3C method and several of its variants as well as the information these methods have revealed about chromatin (see also sidebar Experimental Suitability of the Methods for Long-Range Chromatin Interactions).

3C (Chromosome Conformation Capture)

The main steps of 3C include treating cells with formaldehyde to cross-link in vivo protein-protein and protein-DNA interactions, digesting chromatin with a restriction enzyme, ligating digested chromatin under conditions that favor ligation of fragments in very close

proximity to one another (intramolecular ligation), reversing cross-links, and determining by quantitative polymerase chain reaction (PCR) the frequency of ligation events between genomic regions of interest using sequence-specific primers (**Figure 3*a***, left side).

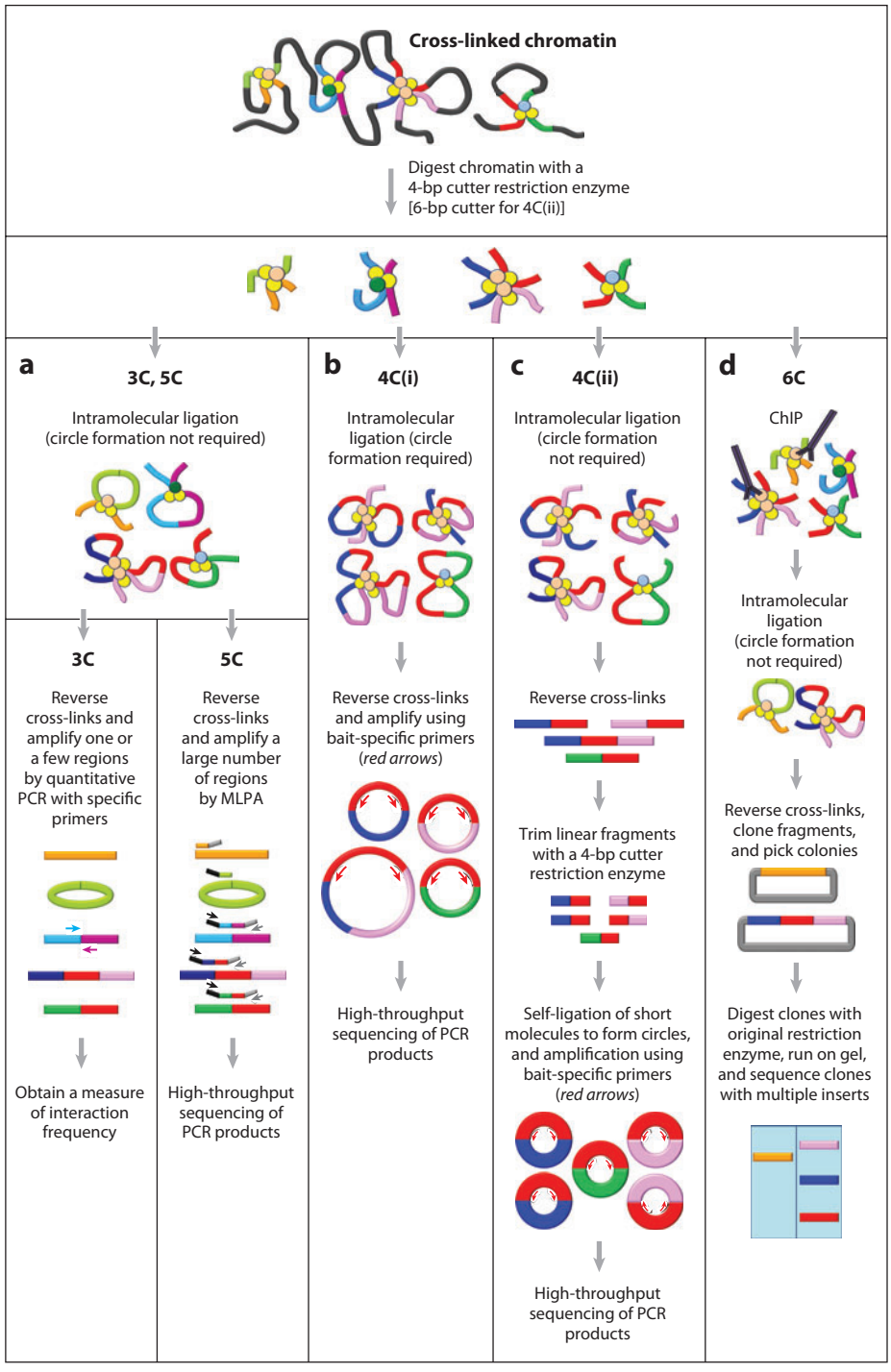
After its initial application in yeast (as mentioned above), 3C was next used to study the chromatin structure of the mouse beta-globin genomic region of approximately 200 kb in both erythroid tissue (where the beta-globin genes are actively transcribed) and brain tissue (where the genes are inactive) (79). The beta-globin genes are developmentally regulated by a locus control region approximately 50 kb away. This region was shown to be in close proximity with the active genes in erythroid but not brain tissue. The intervening chromatin containing the inactive genes looped out. Further experiments with other segments of this genomic region showed that it adopts an erythroid-specific chromatin structure referred to as the active chromatin hub (79). In a later study, it became clear that the looping structure is dynamic even in erythroid cells at different developmental stages, depending on which beta-globin genes are active (60). More recently, 3C was used to show that the region surrounding the rs378854 SNP associated with prostate cancer interacts with the promoters (located 500 kb away) of *PVT1* and *MYC* oncogenes. The presence of the at-risk G allele of this SNP caused an increase in *PVT1* (but not *MYC*) expression in normal prostate tissue, suggesting a functional role for the variant (53).

3C can detect interactions between regions located several thousand to several hundred thousand base pairs away. A variation of the original 3C method substitutes restriction enzyme digestion with a sonication step for fragmenting chromatin. This supposedly generates smaller fragments that are more likely to undergo intramolecular ligation compared with longer fragments, which can encounter other molecules and ligate with them more readily in solution (24). It is also possible to incorporate a ChIP step into the 3C method prior to the ligation step and enrich for fragments bound only by a particular protein of interest (24).

4C (Circular Chromosome Conformation Capture and Chromosome-Conformation-Capture-on-Chip)

The 4C method allows the unbiased detection of all genomic regions that interact with a particular region of interest, referred to as the bait. There are two variants of this method. In the first, following restriction digestion as in the 3C method, ligation is carried out to form a circle such that both ends of the bait are ligated to both ends of any interacting fragment (91). In this method, 4C stands for circular chromosome conformation capture, or simply circular 3C (**Figure 3*b***). Cross-links are then reversed, and linear PCR with bait-specific primers facing outward is used to amplify the interacting fragments, which can be identified using high-density oligonucleotide microarrays or high-throughput sequencing.

In the second variant, 4C stands for chromosome-conformation-capture-on-chip, or simply 3C-on-chip. A 6-bp cutter is used as the restriction enzyme to digest cross-linked chromatin, and ligation is then carried out such that only one end of the bait fragment is required to ligate with one end of the interacting fragment to produce a linear molecule (**Figure 3*c***). Cross-links are then reversed, and the ligation junction of the linear molecule is trimmed by digestion with a 4-bp cutter restriction enzyme (a more frequent cutter than a 6-bp cutter) (74). The trimmed molecule is finally made to undergo self-ligation to produce circles, which are analyzed in the same way as in the circular 3C variant. The 3C-on-chip variant was first used to show that the beta-globin locus in the fetal liver, where it is actively transcribed, made contacts preferentially with other active regions on the same chromosome; in the brain, where it is not transcribed, it made contacts with transcriptionally silent regions (74). More recently, application of 4C revealed that clusters of tRNA genes on human



chromosome 17 (known as tDNA), which are moderately repetitive, act as insulators and make physical contacts with each other but not with promoters of nearby genes (65).

5C (Chromosome Conformation Capture Carbon Copy)

The 5C method is a high-throughput version of 3C. It entails the same steps as 3C, with one main difference in the last step of identifying the interacting fragments. Specifically, multiple PCR primer pairs are used, each containing different genomic sequences of interest being tested for interactions with one another, as well as universal sequences on their 5′ ends (**Figure 3*a***, right side). The primer mix is combined with the 3C library to allow the annealing of only a select number of primer pairs that span the ligation junctions of the 3C library fragments (17). The annealed primers are ligated to each other and amplified using universal primer pairs [multiplex ligation-dependent probe amplification (MLPA)]. Thus, only ligated primer pairs, reflecting a portion of the 3C library containing the interacting fragments, are amplified (17). The PCR products are then identified using microarrays or high-throughput sequencing.

Application of 5C, in combination with a computational integrative modeling platform, to the 500-kb-long genomic region containing human alpha-globin genes helped reveal the first 3D structure of this region (4). In particular, the region was found to exist in a single globular conformation in cells that expressed the alpha-globin genes, but in a double-globule conformation in cells that did not express these genes (but that did express other genes in this region). Furthermore, the inner regions of the globules contained actively transcribed genes, whereas the outer regions were enriched for silent genes (4). This led to the proposal that globules may represent a

Figure 3

Outline of the 3C (chromosome conformation capture) technique and its variants for the detection of long-range chromatin interactions. Cross-linked chromatin is shown at the top with four loops. Interacting chromatin regions are shown in colors other than black. The red fragment is shown twice, representing two different cell types or experimental time points. Although chromatin immunoprecipitation (ChIP) is normally included in 6C (combined chromosome conformation capture ChIP cloning), it is also possible to include this step in all of the other methods. (*a*) The 3C library can be used to carry out both 3C (*left*) and 5C (chromosome conformation capture carbon copy) (*right*). In 3C, only one or a handful of interactions are detected, and their frequencies are determined by quantitative polymerase chain reaction (PCR) with sequence-specific primers. In 5C, multiple interactions are detected by multiplex ligation-dependent probe amplification (MLPA) using a collection of primer pairs containing universal linker sequences. For 5C, one self-ligated (*light green*) and one nonligated (*orange*) product are shown, and the primers complementary to those particular fragments will not ligate and amplify. (*b*) In the 4C(i) method (circular chromosome conformation capture, or circular 3C), only interactions with the red fragment (the bait) are shown. Three possible ligation products of the loop containing three fragments are shown. 4C(i) has fewer processing steps than 4C(ii) but also the disadvantage of requiring circle formation during ligation. This may not be as efficient, especially because it is carried out before cross-link reversal. Also, if the bait interacts with multiple fragments in the same cell type, then the resulting circles will be large and their amplification using bait-specific primers may not be as efficient. (*c*) The 4C(ii) method (chromosome-conformation-capture-on-chip, or 3C-on-chip) has more steps than 4C(i) but does not require circle formation in the initial ligation step. Following cross-link reversal, the linear molecules are longer [because chromatin is initially digested with a 6-base-pair (bp) cutter restriction enzyme] and are trimmed with a 4-bp cutter restriction enzyme to produce shorter fragments. The shorter fragments are then circularized by self-ligation and amplified using bait-specific primers facing outward. Because circles are shorter, it is more efficient to amplify many interacting fragments. (*d*) In 6C, there is a ChIP step after the chromatin has been digested, followed by ligation under dilute conditions and cloning of linear fragments in a vector (*gray*). Self-ligation can also occur here; this is again shown for the light green fragment, which circularizes and does not clone in the vector. The orange fragment remains linear and therefore becomes cloned. After cloning and transforming in bacteria, individual colonies are picked and plasmids are isolated and digested with the same restriction enzyme that was used to digest the chromatin in the first step. The digestion products are then resolved by gel electrophoresis (the left lane is from a clone with a single fragment that did not ligate; the right lane is from a clone with multiple inserts representing interacting segments). Clones containing multiple fragments represent ligation between distal genomic regions that interact.

higher-order chromatin structure where active genes utilizing common/shared transcription machineries cluster.

6C (Combined Chromosome Conformation Capture ChIP Cloning)

The 6C method is essentially the same as 3C, except that after the initial chromosome digestion step it includes a ChIP step followed by intramolecular ligation and cloning of the ChIP-isolated DNA into a vector and transformation in bacteria. The clones showing multiple fragments after digestion with the same restriction enzyme used to build the 3C library are then sequenced from both ends of the vector to identify the interacting partners (78) (**Figure 3*d***).

6C was first implemented in a study that carried out ChIP against the Polycomb-group histone methyltransferase protein EZH2. The study found 5 clones (out of 352) with multiple fragments after digestion, demonstrating that EZH2 was involved in mediating both intra- and interchromosomal long-range interactions that did not occur when EZH2 was knocked down (78). Nevertheless, although the method is feasible, it is labor intensive and not genome-wide because it involves screening individual clones by restriction digestion.

Global 3C Interactions (Hi-C)

The above-mentioned methods rely largely on some prior knowledge of interaction sites, such as promoters of interest, and require large-scale design of oligonucleotides for testing numerous sites. However, this does not allow for the discovery of completely novel interactions, or for an unbiased survey of the genome. To this end, two nearly identical methods were recently developed to identify global looping interactions in yeast and humans in an unbiased manner (18, 47). The first method is known as Hi-C; the second does not have an official name.

Briefly, for Hi-C in humans, cells are cross-linked with formaldehyde; chromatin is digested with a restriction enzyme that produces 5′ overhangs in the DNA; the overhangs are filled in with nucleotides, one of which is biotinylated; ligation is carried out after chromatin dilution to favor intramolecular ligation events; the ligated fragments are sheared; and fragments with biotin at their ligation junctions are isolated by streptavidin and identified by high-throughput sequencing (47) (**Figure 4*a***). In yeast, the method is similar but uses a series of restriction digestion and circularization steps as in 4C and couples this to next-generation sequencing (18). A biotinylation step is also included to allow for selection and enrichment of intramolecular ligation events. In essence, this method involves sampling a much larger proportion of the 3C library generated after the initial digestion of chromatin, thereby allowing for the detection of many more chromatin interactions. In yeast, the small size of the genome provided a higher-resolution map, which revealed what the authors described as a "water lily" structure, with chromosomal arms extending out from a cluster of centromeres (18).

Application of Hi-C in humans showed that open and closed chromatin regions are located in different compartments of the nucleus. Also, at megabase scales, chromatin is organized as a fractal globule—that is, a knot-free polymer that can be readily and densely packaged while at the same time allowing unfolding to occur at any region when needed (47). Thus, the Hi-C data did not support a previously proposed model called the equilibrium globule, in which chromatin exists as a compact and highly knotted polymer.

ChIA-PET (Chromatin Interaction Analysis by Paired-End Tag Sequencing)

The ChIA-PET method, like Hi-C, allows the detection of long-range chromatin interactions on a genome-wide scale in an unbiased fashion. The main steps include cross-linking and sonicating chromatin; carrying out ChIP to enrich for fragments bound by a particular protein of interest; ligating biotinylated linkers (containing an *Mme1* restriction site) to fragment ends, which are then allowed to ligate with

one another under dilute conditions; reversing cross-links; digesting with the *Mme1* restriction enzyme, which cuts 20 bp downstream of its restriction site; isolating fragments with ligation junctions by capturing them with streptavidin; and identifying the fragments by paired-end sequencing (23) (**Figure 4*b***). The fragments therefore contain the two ligated linkers in the middle flanked by genomic DNA sequences on either side. When the tags are mapped to the reference genome, self-ligation events (i.e., those between two ends of the same molecule) are represented by tags less than 3,000 bp from each other. These reflect protein binding sites that would be found by carrying out a standard ChIP experiment. Ligation events between different fragments within the same complex will be represented by tags further apart on the same or even different chromosomes, reflecting long-range interactions (23). Including a ChIP step for specific proteins reduces the complexity of the chromatin interactome map compared with global 3C methods.

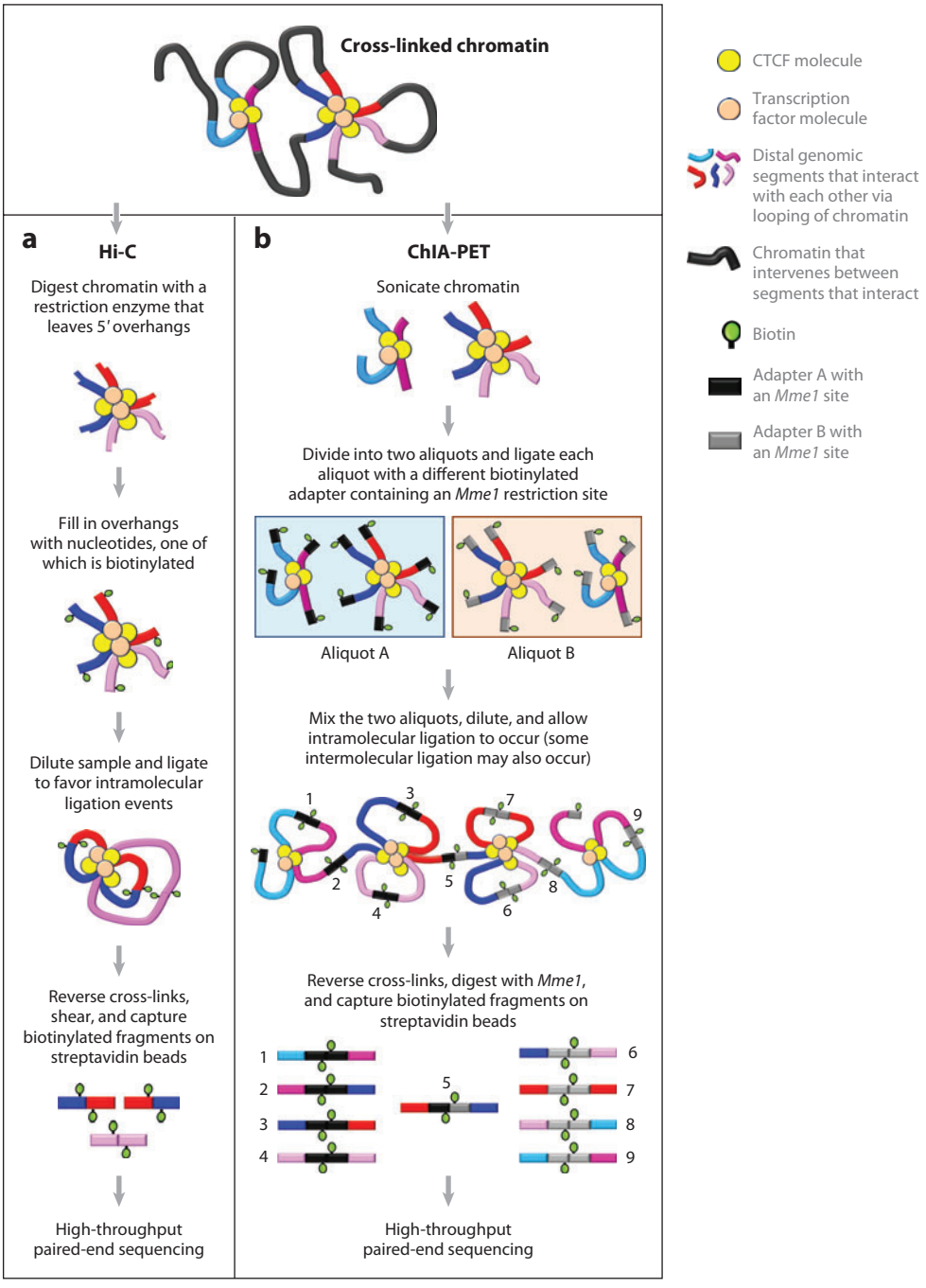
ChIA-PET was first applied to build a chromatin interactome map of estrogen receptor alpha in the estrogen-treated human MCF-7 breast cancer cell line (23). More than 1,000 duplex interactions were found, representing two anchor/interacting regions with a loop between them, almost one-quarter of which turned out to be interconnected because they have a shared anchor (23). These were termed complex interactions. Moreover, genes that were located within 20 kb of anchor regions were more transcriptionally active than those within the loop regions. Thus, there was an association between chromatin structure and gene regulation.

More recently, ChIA-PET has been used to identify with high confidence 1,480 intrachromosomal and 336 interchromosomal interactions involving a total of 3,306 CTCF binding sites in mouse embryonic stem cells (30). By overlaying ChIP data from seven different histone H3 modification marks both inside and outside the CTCF-mediated loops along with information regarding p300 enhancer binding and RNA polymerase II activity, the authors found that the looped chromatin could be categorized into five distinct domains. Each of these domains contained a unique combination of transcriptionally active or inactive genes inside or outside the loops, with one category not having any genes or histone-modification marks (30). Additionally, even though CTCF is classically known for functioning as an insulator binding protein and therefore preventing communication between enhancers and promoters, it can also mediate the interaction of enhancers with promoters of some genes to activate transcription (30). It also facilitates the organization of subnuclear domains by demarcating interactions of chromatin with the nuclear lamina. An intriguing observation is that this study's application of ChIA-PET was able to detect looping interactions between only approximately 10% of all identified CTCF binding sites (14, 30, 43). Given that CTCF functions by binding to distal regulatory elements (such as insulators), one would expect a large fraction of these sites to be involved in loop formation. One explanation is that ChIA-PET may not be a highly sensitive technique for detecting loops. Another is that, owing to inherent biological differences, different numbers of binding sites are involved in loop formation in a given cell type at a given time point. Additionally, CTCF-mediated loop formation likely depends on other factors, such as the cohesin complex, whose chromatin interaction maps would undoubtedly provide additional insights on CTCF-mediated loops (85).

CONCLUSION

Much progress has been made in the past four decades in unraveling the higher-order structure of chromatin. The discovery of the nucleosome as the repeating unit of chromatin, made possible through the use of electron microscopy and biochemical methods, was a huge stride toward achieving this goal. Application of X-ray crystallography proved instrumental in providing the details of how histones and DNA are arranged inside the nucleosome and in showing that the nucleosomes in the 30-nm chromatin fiber are arranged in a zigzag

pattern as opposed to a solenoid-like pattern. Additionally, in metaphase chromosomes the chromatin exists in platelike structures containing interdigitating layers of irregularly oriented nucleosomes. It also appears that not all regions of the genome contain nucleosomes at any given time. Nuclease digestion and other biochemical and genetic methods have



shown that gene regulatory elements generally tend to be nucleosome-depleted, which in turn determines how the overall chromatin is organized. And because different genes are expressed at different times in different cells, chromatin structure is clearly highly dynamic. The advent and use of new genomic technologies have driven our current understanding of nucleosome positioning and dynamics, which has provided insight on gene regulation and the global 3D architecture of chromatin.

Methods such as ChIP and 3C and its variants followed by high-throughput sequencing have resulted in an unprecedented wealth of data that is beginning to shed light on how chromatin may be organized in different cells at different time points. These methods show that there are many long-range chromatin interactions (between distal regulatory elements) that occur through the formation of loops that somehow come together to form chromatin globules. These globules are thought to interact with one another so that they eventually end up separating chromatin into two compartments—active and inactive.

ChIP-seq, DNase-seq, and MNase-seq have been instrumental in determining global maps of regulatory elements. Given that most of these elements are distal, we will only begin to fully understand their roles in gene regulation by determining their interacting partners, e.g., their target promoters. Methods like Hi-C and ChIA-PET are relatively new to the field of chromatin but have the potential to reveal many different types of long-range chromatin interactions. Thus, widespread use of these techniques in the future by many researchers is necessary to determine the dynamics of higher-order chromatin structure. This will be aided in part by a continued decrease in next-generation sequencing cost and the coupled increase of data output.

Figure 4

Outline of Hi-C and ChIA-PET (chromatin interaction analysis by paired-end tag sequencing) methods for genome-wide detection of long-range interactions. Two cross-linked chromatin loops are shown at the top. (*a*) For Hi-C, only one loop is shown for illustration, and a possible self-ligation event is depicted as occurring in the pink fragment. All ligation junctions will contain biotin, and these are captured by using streptavidin beads. Following sequencing, the two ends that map far away from each other in the genome (at least several thousand base pairs or on different chromosomes) represent long-range interactions. Those that map close to each other (several hundred base pairs) represent self-ligation events such as the one shown for the pink fragment. (*b*) For ChIA-PET, two loops are shown. Chromatin is divided into two identical aliquots following sonication. The ends of molecules in each aliquot are ligated to biotinylated adapters that are 16 base pairs (bp) long [adapter A (*shown in black*) for aliquot A, and adapter B (*shown in gray*) for aliquot B]. The adapters contain a restriction site for *Mme1*, which cuts 20 bp away from it. The two aliquots are mixed, and ligation is further carried out under dilute conditions to favor ligation between interacting fragments, that is, intramolecular ligation. Nevertheless, some intermolecular ligation can also occur (shown by ligation junctions 2, 5, and 8); its frequency is determined by the occurrence of hybrid adapters after sequencing (ligation junction 5). Digestion with *Mme1* is carried out to release molecules that contain two adapters in the middle flanked by 20 bp of sequence of the interacting fragments. These are captured by streptavidin beads and sequenced. After sequencing, the reads containing hybrid adapters are counted to ensure that the number is much lower than the number of reads with nonhybrid adapters. Thus, the purpose of dividing chromatin into two aliquots is to provide a quantitative measure of intermolecular ligation, making ChIA-PET a superior method. Note that some nonhybrid adapters also represent intermolecular ligation events (ligation junctions 2 and 8). But because these are ligation events between randomly colliding molecules and do not represent true in vivo long-range interactions, they are observed as paired reads mapping to the genome at a low coverage. Self-ligation events (junctions 4 and 7) are revealed by paired-end reads mapping close to each other in the genome (several hundred base pairs apart). These denote protein binding sites, which can also be found simply by carrying out a standard ChIP. Ligation events between true in vivo distal interacting fragments (intramolecular ligation junctions 1, 3, 6, and 9) are revealed by paired-end reads that map far away from each other in the genome with a high coverage (at least several thousand base pairs apart or on different chromosomes).

Furthermore, development of new computational tools able to integrate and interpret the data sets generated by these methods is also needed. There is already a tool to automatically analyze ChIA-PET data (ChIA-PET Tool, available at **http://chiapet.gis.a-star.edu.sg**) (46). Another example is the integrative modeling platform tool, which was used to construct a 3D model of chromatin based on interactions in the 500-kb-long human alpha-globin genomic region (4). Further developments in 3D chromatin modeling and polymer models such as the fractal globule model mentioned above will pave the way for a more comprehensive view of higher-order chromatin structure.

SUMMARY POINTS

1. The recent advent of high-throughput genomic techniques such as high-density microarrays, and particularly next-generation high-throughput sequencing, has been critical in assessing chromatin structure on a large scale in a short time.
2. MNase digestion has helped determine patterns of nucleosome occupancy, whereas DNaseI digestion has helped determine genomic regions depleted in nucleosomes.
3. ChIP has facilitated the isolation and examination of only those regions of the genome that are bound by a protein of interest.
4. Loop formation is an integral component of chromatin packaging made possible through protein-mediated interactions between genomic regions far apart from one another (such as promoters and enhancers). These interactions can be detected by recently developed methods such as 3C and several of its variants.
5. It is speculated that many loops, formed as a result of several genes being transcribed, interact with one another to form chromatin globules. These go on to further interact with one another to subsequently divide the chromatin into two large blocks of active and inactive domains.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

1. Arya G, Maitra A, Grigoryev SA. 2010. A structural perspective on the where, how, why, and what of nucleosome positioning. *J. Biomol. Struct. Dyn.* 27:803–20
2. Auerbach RK, Euskirchen G, Rozowsky J, Lamarre-Vincent N, Moqtaderi Z, et al. 2009. Mapping accessible chromatin regions using Sono-Seq. *Proc. Natl. Acad. Sci. USA* 106:14926–31
3. Basehoar AD, Zanton SJ, Pugh BF. 2004. Identification and distinct regulation of yeast TATA box-containing genes. *Cell* 116:699–709
4. Bau D, Sanyal A, Lajoie BR, Capriotti E, Byron M, et al. 2011. The three-dimensional folding of the α-globin gene domain reveals formation of chromatin globules. *Nat. Struct. Mol. Biol.* 18:107–14
5. Bhaumik SR, Smith E, Shilatifard A. 2007. Covalent modifications of histones during development and disease pathogenesis. *Nat. Struct. Mol. Biol.* 14:1008–16
6. **Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, et al. 2008. High-resolution mapping and characterization of open chromatin across the genome. *Cell* 132:311–22**
7. Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, et al. 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.* 18:630–34

6. The first study to employ both DNase-chip and true high-throughput sequencing to identify DHSs on a genome-wide scale.

8. Caravaca JM, Cano S, Gallego I, Daban JR. 2005. Structural elements of bulk chromatin within metaphase chromosomes. *Chromosome Res.* 13:725–43
9. Castro-Hartmann P, Milla M, Daban JR. 2010. Irregular orientation of nucleosomes in the well-defined chromatin plates of metaphase chromosomes. *Biochemistry* 49:4043–50
10. Chodavarapu RK, Feng S, Bernatavichute YV, Chen PY, Stroud H, et al. 2010. Relationship between nucleosome positioning and DNA methylation. *Nature* 466:388–92
11. Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, et al. 2008. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* 452:215–19
12. Crawford GE, Davis S, Scacheri PC, Renaud G, Halawi MJ, et al. 2006. DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. *Nat. Methods* 3:503–9
13. Crawford GE, Holt IE, Whittle J, Webb BD, Tai D, et al. 2006. Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). *Genome Res.* 16:123–31
14. Cuddapah S, Jothi R, Schones DE, Roh TY, Cui K, Zhao K. 2009. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res.* 19:24–32
15. Defossez PA, Stancheva I. 2011. Biological functions of methyl-CpG-binding proteins. *Prog. Mol. Biol. Transl. Sci.* 101:377–98
16. Dekker J, Rippe K, Dekker M, Kleckner N. 2002. Capturing chromosome conformation. *Science* 295:1306–11
17. Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, et al. 2006. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* 16:1299–309
18. Duan Z, Andronescu M, Schutz K, McIlwain S, Kim YJ, et al. 2010. A three-dimensional model of the yeast genome. *Nature* 465:363–67
19. Earnshaw WC, Laemmli UK. 1983. Architecture of metaphase chromosomes and chromosome scaffolds. *J. Cell Biol.* 96:84–93
20. Farnham PJ. 2009. Insights from genomic profiling of transcription factors. *Nat. Rev. Genet.* 10:605–16
21. Finch JT, Lutter LC, Rhodes D, Brown RS, Rushton B, et al. 1977. Structure of nucleosome core particles of chromatin. *Nature* 269:29–36
22. Fu Y, Sinha M, Peterson CL, Weng Z. 2008. The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet.* 4:e1000138
23. Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, et al. 2009. An oestrogen-receptor-α-bound human chromatin interactome. *Nature* 462:58–64
24. Fullwood MJ, Ruan Y. 2009. ChIP-based methods for the identification of long-range chromatin interactions. *J. Cell. Biochem.* 107:30–39
25. Gabdank I, Barash D, Trifonov EN. 2009. Nucleosome DNA bendability matrix (*C. elegans*). *J. Biomol. Struct. Dyn.* 26:403–11
26. Gallego I, Castro-Hartmann P, Caravaca JM, Cano S, Daban JR. 2009. Dense chromatin plates in metaphase chromosomes. *Eur. Biophys. J.* 38:503–22
27. Gaulton KJ, Nammo T, Pasquali L, Simon JM, Giresi PG, et al. 2010. A map of open chromatin in human pancreatic islets. *Nat. Genet.* 42:255–59
28. Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res.* 17:877–85
29. Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, et al. 2006. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat. Genet.* 38:320–23
30. Handoko L, Xu H, Li G, Ngan CY, Chew E, et al. 2011. CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat. Genet.* 43:630–38
31. Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, et al. 2010. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* 6:479–91
32. Hawkins RD, Hon GC, Yang C, Antosiewicz-Bourget JE, Lee LK, et al. 2011. Dynamic chromatin states in human ES cells reveal potential regulatory sequences and genes involved in pluripotency. *Cell Res.* 21:1393–409

12. One of the first two studies to interrogate DHSs on a genome-wide scale using DNase-chip.

16. Introduces 3C to detect long-range chromatin interactions and shows that yeast chromosome III acquires a contorted ring conformation.

23. Uses ChIA-PET to show that only 10% of all CTCF binding sites are actually involved in loop formation.

30. Introduces ChIA-PET and shows its feasibility by determining that long-range interactions can be mediated by estrogen receptor alpha.

33. He A, Pu WT. 2010. Genome-wide location analysis by pull down of in vivo biotinylated transcription factors. *Curr. Protoc. Mol. Biol.* 92:21.20.1–15
34. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, et al. 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459:108–12
35. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, et al. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* 39:311–18
36. Helgason A, Palsson S, Thorleifsson G, Grant SF, Emilsson V, et al. 2007. Refining the impact of TCF7L2 gene variants on type 2 diabetes and adaptive evolution. *Nat. Genet.* 39:218–25
37. Henikoff JG, Belsky JA, Krassovsky K, Macalpine DM, Henikoff S. 2011. Epigenome characterization at single base-pair resolution. *Proc. Natl. Acad. Sci. USA* 108:18318–23
38. Hewish DR, Burgoyne LA. 1973. Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. *Biochem. Biophys. Res. Commun.* 52:504–10
39. Ioshikhes IP, Albert I, Zanton SJ, Pugh BF. 2006. Nucleosome positions predicted through comparative genomics. *Nat. Genet.* 38:1210–15
40. Johnson SM, Tan FJ, McCullough HL, Riordan DP, Fire AZ. 2006. Flexibility and constraint in the nucleosome core landscape of *Caenorhabditis elegans* chromatin. *Genome Res.* 16:1505–16
41. Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, et al. 2009. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* 458:362–66
42. Keene MA, Corces V, Lowenhaupt K, Elgin SC. 1981. DNase I hypersensitive sites in *Drosophila* chromatin occur at the 5′ ends of regions of transcription. *Proc. Natl. Acad. Sci. USA* 78:143–46
43. Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, et al. 2007. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* 128:1231–45
44. Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J. 2009. Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat. Genet.* 41:376–81
45. Kornberg RD. 1974. Chromatin structure: a repeating unit of histones and DNA. *Science* 184:868–71
46. Li G, Fullwood MJ, Xu H, Mulawadi FH, Velkov S, et al. 2010. ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. *Genome Biol.* 11:R22
47. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326:289–93
48. Lowary PT, Widom J. 1998. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* 276:19–42
49. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* 389:251–60
50. Magnani L, Ballantyne EB, Zhang X, Lupien M. 2011. PBX1 genomic pioneer function drives ERα signaling underlying progression in breast cancer. *PLoS Genet.* 7:e1002368
51. Mavrich TN, Ioshikhes IP, Venters BJ, Jiang C, Tomsho LP, et al. 2008. A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res.* 18:1073–83
52. Mavrich TN, Jiang C, Ioshikhes IP, Li X, Venters BJ, et al. 2008. Nucleosome organization in the *Drosophila* genome. *Nature* 453:358–62
53. Meyer KB, Maia AT, O'Reilly M, Ghoussaini M, Prathalingam R, et al. 2011. A functional variant at a prostate cancer predisposition locus at 8q24 is associated with *PVT1* expression. *PLoS Genet.* 7:e1002165
54. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, et al. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448:553–60
55. Nagy PL, Cleary ML, Brown PO, Lieb JD. 2003. Genomewide demarcation of RNA polymerase II transcription units revealed by physical fractionation of chromatin. *Proc. Natl. Acad. Sci. USA* 100:6364–69
56. Nammo T, Rodriguez-Segui SA, Ferrer J. 2011. Mapping open chromatin with formaldehyde-assisted isolation of regulatory elements. *Methods Mol. Biol.* 791:287–96
57. Noll M, Kornberg RD. 1977. Action of micrococcal nuclease on chromatin and the location of histone H1. *J. Mol. Biol.* 109:393–404
58. Olins AL, Olins DE. 1974. Spheroid chromatin units (v bodies). *Science* 183:330–32

47. Describes Hi-C and determines that at the megabase scale chromatin exists as a fractal globule, that is, a knot-free polymer.

59. Oudet P, Gross-Bellard M, Chambon P. 1975. Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell* 4:281–300
60. Palstra RJ, Tolhuis B, Splinter E, Nijmeijer R, Grosveld F, de Laat W. 2003. The β-globin nuclear compartment in development and erythroid differentiation. *Nat. Genet.* 35:190–94
61. Pan G, Tian S, Nie J, Yang C, Ruotti V, et al. 2007. Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell* 1:299–312
62. Park PJ. 2009. ChIP-seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.* 10:669–80
63. Pauli U, Chrysogelos S, Nick H, Stein G, Stein J. 1989. In vivo protein binding sites and nuclease hypersensitivity in the promoter region of a cell cycle regulated human H3 histone gene. *Nucleic Acids Res.* 17:2333–50
64. Paulson JR, Laemmli UK. 1977. The structure of histone-depleted metaphase chromosomes. *Cell* 12:817–28
65. Raab JR, Chiu J, Zhu J, Katzman S, Kurukuti S, et al. 2011. Human tRNA genes function as chromatin insulators. *EMBO J.* 31:330–50
66. Richmond TJ, Davey CA. 2003. The structure of DNA in the nucleosome core. *Nature* 423:145–50
67. Robinson PJ, Fairall L, Huynh VA, Rhodes D. 2006. EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure. *Proc. Natl. Acad. Sci. USA* 103:6506–11
68. Sabo PJ, Kuehn MS, Thurman R, Johnson BE, Johnson EM, et al. 2006. Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nat. Methods* 3:511–18
69. Schalch T, Duda S, Sargent DF, Richmond TJ. 2005. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* 436:138–41
70. Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, et al. 2008. Dynamic regulation of nucleosome positioning in the human genome. *Cell* 132:887–98
71. Segal E, Fondufe-Mittendorf Y, Chen L, Thastrom A, Field Y, et al. 2006. A genomic code for nucleosome positioning. *Nature* 442:772–78
72. Shivaswamy S, Bhinge A, Zhao Y, Jones S, Hirst M, Iyer VR. 2008. Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. *PLoS Biol.* 6:e65
73. Shogren-Knaak M, Ishii H, Sun JM, Pazin MJ, Davie JR, Peterson CL. 2006. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* 311:844–47
74. Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, et al. 2006. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.* 38:1348–54
75. Singh BN, Hampsey M. 2007. A transcription-independent role for TFIIB in gene looping. *Mol. Cell* 27:806–16
76. Song L, Zhang Z, Grasfeder LL, Boyle AP, Giresi PG, et al. 2011. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res.* 21:1757–67
77. Thomas S, Li XY, Sabo PJ, Sandstrom R, Thurman RE, et al. 2011. Dynamic reprogramming of chromatin accessibility during *Drosophila* embryo development. *Genome Biol.* 12:R43
78. Tiwari VK, Cope L, McGarvey KM, Ohm JE, Baylin SB. 2008. A novel 6C assay uncovers Polycomb-mediated higher order chromatin conformations. *Genome Res.* 18:1171–79
79. Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W. 2002. Looping and interaction between hypersensitive sites in the active β-globin locus. *Mol. Cell* 10:1453–65
80. Towbin BD, Meister P, Gasser SM. 2009. The nuclear envelope—a scaffold for silencing? *Curr. Opin. Genet. Dev.* 19:180–86
81. Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, et al. 2008. A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Res.* 18:1051–63
82. Waki H, Nakamura M, Yamauchi T, Wakabayashi K, Yu J, et al. 2011. Global mapping of cell type–specific open chromatin by FAIRE-seq reveals the regulatory role of the NFI family in adipocyte differentiation. *PLoS Genet.* 7:e1002311
83. Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, et al. 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* 40:897–903

68. One of the first two studies to interrogate DHSs on a genome-wide scale using DNase-array.

84. Weintraub H, Groudine M. 1976. Chromosomal subunits in active genes have an altered conformation. *Science* 193:848–56
85. Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, et al. 2008. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 451:796–801
86. Wu C, Bingham PM, Livak KJ, Holmgren R, Elgin SC. 1979. The chromatin structure of specific genes: I. Evidence for higher order domains of defined DNA sequence. *Cell* 16:797–806
87. Wu C, Wong YC, Elgin SC. 1979. The chromatin structure of specific genes: II. Disruption of chromatin structure during gene activity. *Cell* 16:807–14
88. Xi H, Shulha HP, Lin JM, Vales TR, Fu Y, et al. 2007. Identification and characterization of cell type–specific and ubiquitous chromatin regulatory structures in the human genome. *PLoS Genet.* 3:e136
89. Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, et al. 2005. Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* 309:626–30
90. Zhang Y, Shin H, Song JS, Lei Y, Liu XS. 2008. Identifying positioned nucleosomes with epigenetic marks in human from ChIP-Seq. *BMC Genomics* 9:537
91. Zhao Z, Tavoosidana G, Sjolinder M, Gondor A, Mariano P, et al. 2006. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.* 38:1341–47
92. Zhou VW, Goren A, Bernstein BE. 2011. Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.* 12:7–18

Annual Review of Genomics and Human Genetics

Volume 13, 2012

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Genomics and Human Genetics* articles may be found at http://genom.annualreviews.org/errata.shtml