Review TRENDS in Genetics Vol.23 No.2

Chromatin structure in the genomics era

Oliver J. Rando

Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, USA

The packaging of eukaryotic genomes into chromatin has a large influence on DNA-templated processes, such as transcription. The availability of genome sequences and 'genomics' technologies such as DNA microarrays and high-throughput sequencing had an immediate effect on the study of transcriptional regulation, by enabling researchers to identify the coregulation patterns of thousands of genes. These same resources are now being used successfully to study the structure of chromatin. Here, I review some of these new genomics approaches to understanding chromatin structure in eukaryotes.

Introduction

Eukaryotic DNA is packaged into a nucleoprotein complex known as chromatin, and this packaging has major functional consequences for most processes that involve DNA. Understanding processes such as DNA damage, transcriptional signal processing and cellular differentiation probably requires a detailed understanding of the chromatin context of the genome. The 'genomics era' has yielded great insights in other areas, with genome sequencing yielding extensive information about genomic organization and evolution, and with microarray measurements of transcript abundance considerably improving our understanding of transcriptional control and signal processing. The purpose of this article is to summarize the insights that have been gained by applying genomics approaches such as microarrays and high-throughput sequencing to chromatin structure.

Chromatin structure

The positioning and modification state of nucleosomes influence processes from transcription to DNA repair to replication timing. In addition to its roles in plastic responses to the environment, chromatin seems to be capable of carrying epigenetic information for many generations [1,2]. Similarly, many eukaryotes methylate a subset of cytosine bases in their genome, and this methylation regulates various DNA-templated processes (e.g. transposon silencing and telomere length maintenance) and provides a carrier for epigenetic information [3]. Both cytosine methylation and the histones are thought to be localized to a particular genomic location for longer time scales than typical DNA-binding proteins such as transcription factors (although counterexamples can be found) and, therefore, are often considered in a structural context.

As noted in Box 1, chromatin organization can conveniently be considered by analogy to protein folding. Here, I describe genomics approaches to characterizing chromatin at the levels of primary and secondary structure, and I summarize selected insights gained from these studies. The following related topics might be of interest: genomic localization of DNA-binding proteins such as transcription factors [4–6]; light-microscopy imaging of chromatin structure (e.g. in *Drosophila melanogaster* polytene chromosomes and mammalian tissue culture cells) [7,8]; and microarray studies of global gene expression changes in various chromatin mutants [9–11]. This article focuses on studies in which the structural aspects of genome packaging are measured.

Experimental approaches used for chromatin analysis

Most high-resolution genomic localization studies use DNA microarray technology as a read-out (Box 2), although sequencing has also been used as a read-out for certain experiments [12–15] and might become more widespread as high-throughput genome sequencing becomes cheaper and more accessible to most researchers [16]. Sequencing methods are widely understood: a sample of interest is prepared – by immunoprecipitation (IP), nuclease digestion or chemical treatment of DNA (to study cytosine methylation; discussed later) – and high-depth sequencing is used to characterize the isolated (or modified) DNA populations. To increase throughput, Roh *et al.* [14] used a method named serial analysis of gene expression (SAGE) [17], in which numerous short tags are prepared from isolated DNA and then concatenated to enable each sequencing read to cover numerous distinct isolated fragments.

DNA microarrays are used to characterize the differences between two populations of DNA or RNA. The two main variables in microarray studies are the type of microarray used and the method used for fractionating nucleotides before labeling and microarray hybridization. Early microarray studies were limited by the microarrays that were available, which often consisted of large (~1 kb) PCR products or short (~25–70 bp) oligonucleotides scattered through various coding regions. More recently, several companies have developed tiling microarrays, which have short oligonucleotides spaced at uniform density throughout large regions of the genome, enabling high-resolution localization measurements. Furthermore, many companies now offer

Corresponding author: Rando, O.J. (Oliver.Rando@umassmed.edu).
Available online 22 December 2006.

 0168-9525/$ – see front matter doi:10.1016/j.tig.2006.12.002

Box 1. The elements of chromatin structure

Chromatin is the nucleoprotein packaging of the eukaryotic genome, and its structure can be considered using analogies with protein structure. The primary structure – the 'sequence' – of chromatin consists of a 10 nm fiber, which is observed, when using electron microscopy, as 'beads on a string'. The repeating subunit of the primary structure of chromatin (the 'bead') is the nucleosome, which consists of 147 bp of DNA wrapped around an octamer of basic histone proteins. Nucleosomes at different locations with respect to the underlying primary structure of genomic DNA vary in subunit composition and covalent modification state, and combinations of these nucleosomes can be considered in some ways analogous to the amino acids in a protein. For example, acetylation of lysine residues in the histone tails removes a positive charge that might otherwise mediate internucleosomal interactions, so deacetylated nucleosomes are more likely to occupy positions within particular secondary structural elements. Similarly, proline typically does not occur in α helices in proteins. The number of functionally distinct 'flavors' of nucleosome is unknown. So how important is the difference between a nucleosome carrying H3K4me3K14ac (i.e. histone H3 with Lys4 trimethylated and Lys14 acetylated) and a nucleosome carrying H3K4me3K18ac? Broadly considered, at present, this is a major question in the field of chromatin structure.

The secondary structure of proteins denotes the structural folds that are present: for example, α helices. Analogously, the primary (beads on a string) structure of chromatin compacts *in vitro* into a condensed fiber called the 30 nm fiber. Recent work indicates that the 30 nm fiber consists of the 10 nm fiber folded in a zigzag manner [80]. The buffer dependence of 30-nm-fiber folding *in vitro* raises the possibility that there are other secondary structural elements *in vivo* that have not been detected owing to technical limitations. Chromatin also folds at intermediate scales between the 30 nm fiber and the whole-chromosome fold (i.e. the tertiary structure): for example, the ~80–120 kb looping that occurs between matrix-associated regions.

The tertiary structure of proteins denotes the folding of an entire polypeptide chain. The analogous structure for chromatin is the overall folding of a chromosome.

The quaternary structure of proteins denotes the packing of individual polypeptide chains. By analogy, that of chromatin denotes the orientation of different chromosomes with respect to one another in the nucleus and their positioning relative to markers such as the nuclear periphery.

custom-designed microarrays with relatively low 'up front' costs, thereby considerably lowering the barrier for entry to this area of research.

Considering fractionation methods, the most commonly used is chromatin IP (ChIP), in which formaldehyde-crosslinked chromatin fragments are isolated using antibodies that recognize specific proteins or specific proteins with particular modifications: for example, histone H4 acetylated on Lys8 (H4K8ac). Immunoprecipitated DNA is labeled and hybridized to the microarray (resulting in these studies being termed ChIP–chip studies), revealing regions of the genome that are associated with the protein of interest. Another common technique used to fractionate chromatin is nuclease digestion, because packaging of the genome into chromatin domains affects its accessibility to nucleases. Other fractionation methods that have been described to date rely on the variable solubility of formaldehyde-crosslinked chromatin in organic solvents [18], sucrose gradient fractionation based on the extent of compaction [19], and the deamination of cytosine but not methylcytosine by treatment of genomic DNA with bisulfite [13,15,20,21]. In this article, I describe some biological results obtained using these fractionation techniques.

Box 2. Sequencing versus microarray read-outs for genomic studies

To date, the two main assay technologies that have been used for genomic studies of chromatin are sequencing and microarrays. Both of these technologies can be used to characterize nucleotide populations after isolation, and both have advantages and disadvantages. Microarrays typically are used to compare the relative abundances of a collection of sequences from two samples. (Although single-color hybridization is also a widely used experimental strategy, researchers usually compare two different single-color hybridizations.) Therefore, microarrays provide relative measures of the abundance of a given sequence. Microarray studies are limited by the probe sequences chosen for the microarray, with typical numbers of sequences ranging from ~5000 to several million. Furthermore, although microarrays can distinguish sequences that differ at a single nucleotide, this is still technically challenging and requires custom microarrays. One advantage is that, at present, microarray studies are much cheaper than sequencing studies of a similar scale, and, after a microarray has been built, the time required for an experiment is short.

Although sequencing is expensive compared with microarray studies, it has certain advantages: (i) sequencing provides single-base resolution; even when using densely tiled microarrays, it is difficult to achieve single-base resolution; (ii) when coupled with an appropriate isolation method, sequencing can provide some of the advantages of single-molecule studies (e.g. the variability in the ends of molecules can be characterized); by contrast, microarrays measure the aggregate behavior of a population; (iii) sequencing studies require no assumptions about which sequences will be isolated; by contrast, microarrays are biased by the choice of sequences printed on the array.

Genomics approaches to chromatin primary structure

At least three types of inhomogeneity can affect the primary structure of chromatin: (i) nucleosomes can occlude underlying genome sequences, so the position of a nucleosome along the genome has regulatory consequences [22–25]; (ii) the histone octamer does not have a uniform composition, and alternative isoforms of canonical histones (e.g. the H3 variant H3.3) can be assembled into the octamer in place of the canonical subunit [26]; (iii) histones are subject to an astonishing number and variety of covalent modifications, and nucleosomes at different locations can differ in their pattern of covalent modification [27–29].

Nucleosome positioning and occupancy

In what was perhaps the first microarray study of nucleosome occupancy, differential partitioning of formaldehyde-crosslinked yeast chromatin during phenol–chloroform extraction was used [18]. DNA associated with small amounts of protein is found in the aqueous layer. In this study, intergenic regions were recovered preferentially in the aqueous phase, indicating that these were depleted of nucleosomes [a technique now referred to as formaldehyde-assisted isolation of regulatory elements (FAIRE)]. This was confirmed in two subsequent ChIP–chip studies in which crosslinked chromatin was immunoprecipitated using antibodies specific for histone H3 [30,31]. Low nucleosome occupancy at gene promoters has also been observed in *Drosophila*, from which nucleosomal DNA was isolated by avidin binding to biotinylated histones engineered to carry a recognition sequence for the bacterial biotin ligase BirA [32]. Hogan *et al.* have recently extended