

MBB206 Topic 4 Chromatin

A. Overall structure

The one-dimensional length of the human genome is far greater than the three-dimensional size of the nucleus into which it must fit. To fit into the nucleus the DNA has to be condensed in some manner. The degree to which DNA is condensed is expressed as a **packing ratio**.

Packing ratio: the length of DNA divided by the length into which it is packaged

Here's one example. The shortest human chromosome contains 4.6×10^7 bp of DNA (about 10 times the genome size of *E. coli*). This is equivalent to 14,000 μm of extended DNA. In its most condensed state during mitosis, the chromosome is about 2 μm long. This gives a packing ratio of **7000** (14,000/2). Overall, the entire human genome is about 1 meter in length stretched end-to-end, and it packs into nuclei, the size of $10\mu\text{M}^3$!!!!!! Is this not remarkable?

How eukaryotes achieve a packing ratio ~ 7,000:

To achieve the overall packing ratio of 7000-10,000, DNA is not packaged directly into the final structure we refer to as chromatin. Instead, it contains several hierarchies of organization. The first level of packing is achieved by the winding of DNA around a protein core to produce a "bead-like" structure called a **nucleosome**. This gives a packing ratio of about 6. This structure is basically the same between euchromatin and heterochromatin (see below) but modifications of the proteins in this nucleosome cause important structural differences that we will get into later. The second level of packing is the coiling of beads in a helical structure called the **30 nm fiber** that is found in both interphase chromatin and mitotic chromosomes. This structure increases the packing ratio to about 40. The final packaging occurs when the fiber is organized in **loops, scaffolds and domains** that give a final packing ratio of about 1000 in interphase chromosomes and about 10,000 in mitotic chromosomes.

B. Heterochromatin and Euchromatin

Heterochromatin and Euchromatin refer to particular "states" of chromatin, not any specific set of sequences per se. When chromosomes are stained with dyes, they appear to have alternating lightly and darkly stained regions. The lightly-stained regions are euchromatin and contain single-copy, genetically-active DNA. The darkly-stained regions are heterochromatin and contain repetitive sequences that are "genetically inactive" as well as some genes that are typically inactive. Euchromatin can be considered to contain the majority of genes. Not all of the genes in euchromatin are active, some are actively silenced, and some have a closed type of chromatin structure. Again, these differences are directly linked to presence or absence of DNA and protein modifications at those loci.

Distinctions between Euchromatin and Heterochromatin. From S. Henikoff. 2000 BBA 1470:1-8.

Feature	Euchromatin	Heterochromatin
Interphase appearance	uncondensed	Condensed
Chromosomal location	distal	Pericentromeric and Telomeric
Sequence composition	Mostly non-repetitive	Repetitive
Gene density	high	Low or absent
Replication timing	Throughout S phase	Late S phase
Position effect variegation	rare	Frequent

Methylation state	CpG islands hypomethylated	Fully methylated
Histone acetylation	Low to high	Low
Nucleosome spacing	variable	Regular
Nucleosome accessibility	Variable	Low
Transcription factor binding	variable	Low*

* in general, transcription factors bind poorly to heterochromatin. However, there are proteins that are specifically targeted to heterochromatin and help maintain a repressive environment for gene expression.

C. Nucleosomes

First nucleosome structure solved by Luger et al. in 1997. Nature 389:251-260.
A second nucleosome structure with an alternative H2A histone (H2A.Z) was solved more recently: Suto et al. 2000. Nat Struc Biol. 7:1121-1124.

The nucleosome is the fundamental repeat unit of chromatin and consists of about 146 bp (1.65 turns of the DNA helix) wrapped around a histone octamer. This wrapped structure causes negative supercoiling of the DNA. The core nucleosome contains two copies of core histone, each of which consist of histone proteins: a H2A-H2B dimer and a H3-H4 dimer. Histones are basic proteins that have an affinity for DNA and are the most abundant proteins associated with DNA. The amino acid sequence of each of these four histones is highly conserved suggesting a similar function for all organisms. In general there are five different histones that make up chromosomal DNA:

Nucleosome **core** repeating unit = DNA•[H2A/H2B•H3/H4]₂

Nucleosomes are connected by DNA•H1 or H5 in a **linker** region

There are other subtypes of histones, eg. H2A.Z only some of which are now being studied for unique localization or regulatory properties. H2A.Z appears to be enriched in nucleosomes around actively transcribed genes.

One nucleosome is about 6nm in width. Multiple nucleosomes along a stretch of DNA has about a 10nm width. This "beads-on-a-string" structure can pack down like a slinky, or a spring with about 6 nucleosomes per 360° turn. The width of the slinky or spring is now 30nm. The 30nm fiber can be seen by EM. The 30 nm fiber is then organized into loops and structural domains, presumably by anchorage to nuclear matrix. A similar sort of arrangement is seen with the E. coli genome, where loops are anchored to a support.

D. Histones

can be thought of as having three parts to their structure:

1. **histone-fold motif**

histone folds are comprised of a long alpha helix linked at either end to two shorter alpha helices via short beta turns. There are other proteins that have histone-like folds. We have mentioned several that are in proteins of the TFIID complex.

2. **extension** - this part of each histone is arginine-rich, and therefore positively charged. These regions insert through the tightly packed minor grooves of the wrapped DNA, making extensive contacts.

3. **tails** - at the outside end of the extension are the famous histone "tails". These tails are also rich in basic residues, particularly lysine, as well as serines and threonines. These tails can be

extensively modified by acetylation, methylation, phosphorylation, and even ubiquitination. Each modification may have different consequences and modification of particular residues may be clues to different activities. Clearly, we do not completely understand the full extent of histone tail modifications and their role in gene expression.

Histone tails also make extensive contact with DNA (being positively charged, they particularly like to interact with the negatively charged phosphates). The tail of Histone H4 is interesting: it stretches all the way over and through the DNA of an adjacent nucleosome to contact H2A/H2B dimers. It is interesting then to note that H4 tails are particularly well acetylated during gene activation. Although it is still speculation at this point, most people think of acetylation as breaking electrostatic charge interactions between the histone tail and DNA and loosening contacts between nucleosomes. This loosening of contact would lead to an opening up of the chromatin for transcription.

DNA binding transcription factors cannot bind well to chromatin, K_D 's are increased

Covalent modifications of chromatin and remodeling of the nucleosomes (repositioning) have shown to increase DNA binding by transcription factors and the basic transcription machinery - gene activity is increased.

E. Chromatin modifications

- a. acetylation
- b. phosphorylation
- c. methylation
- d. ubiquitination

acetylation

ϵ -group of the lysine side chain becomes modified by covalent ligation of an acetyl group. This covalent modification neutralizes the positive charge on the amino acid side chain and appears to disrupt electrostatic interaction with the negatively charged DNA wrapped around the nucleosome. The caveat here is that some of the residues in the tail that are modified, don't directly interact with DNA in the first place. Nevertheless, this type of chromatin disruption is thought to be one of the major ways that chromatin is "opened up" for transcription.

Histone Acetyl Transferases - HATs

The first acetylating enzyme identified was by David Allis and his colleagues a little over 10 years ago. Working in *Tetrahymena*, Dr. Allis purified a histone acetylating enzyme activity. It had been known for decades that histones were acetylated, but no enzyme had been identified. Purification and subsequent cloning of this enzyme showed that the amino acid sequence had highly significant homology to a previously identified yeast protein called GCN5. It wasn't known that GCN5 was a HAT - but it was known that GCN4 required this protein to activate gene targets. Thus, the first link between gene activation and chromatin acetylation was made. Mutations of GCN5 that killed HAT activity, destroyed the ability of GCN4 to activate gene transcription and acetylation of proximal nucleosomes was gone.

By now, a large number of HATs have been identified (over 14 in mammalian systems). For a comprehensive list please see:

Lee and Young. 2000. Transcription of Eukaryotic Protein-Coding Genes. *Ann. Rev. Genet.* **34**:77-137
(a very good, up-to-date general review of Pol II transcription)

We will only mention a few notable ones here: GCN5, CBP/p300, NuA4 and TAF_{II}250 (the largest TAF in TFIID)

Purification of these HAT enzymes shows that often (most of the time?) they are part of larger, multi-subunit complexes. For example purification of GCN5 showed that it co-purified and was intrinsically associated with a complex called SAGA (for Spt, Ada, GCN5 Acetyltransferase). Identification of Spt and Ada proteins had already occurred in other yeast studies as proteins essential for transcription start site selection and transcription activation. **Other subunits in SAGA are some of the same ones found in TFIID - the TAFs.**

Why should GCN5 and other HAT's be part of a multi-subunit complex?

- recruitment

many transcription factors can recruit the complex via a large, varied selection of protein•protein interactions

- substrate access

GCN5 can acetylate histones in solution, but not histones within a nucleosome. However, as part of the SAGA complex, GCN5 can easily acetylate specific residues of the H3 and H2B histone tails within the nucleosome.

- substrate specificity

GCN5 acetylates additional lysine residues in a nucleosome when it is part of the SAGA complex. Also specificities may be restricted, different HAT complexes acetylate different residues on different histones

Purification of HAT activity using an *in vitro* histone acetylation assay led to the realization that other HATs are also part of large multi-subunit complexes.

- GCN5 is part of SAGA
- TAF_{II}250 is part of TFIID
- TFIIC contains a HAT
- CBP/p300 - although usually thought of as a single entity - it binds to other co-activator complexes (which themselves have HAT activity such as P/CAF) and enables their recruitment to active genes.

Most or all of these complexes are recruited by interactions with the activation domains of sequence-specific DNA binding proteins.

Phosphorylation

Phosphorylation of serine residues of the tails of H1 and H3 are found enriched in condensing chromatin during mitosis. In addition, David Allis has found that phosphorylation of a particular lysine residue of H3 correlates extremely well with actively transcribed genes. Antibodies specific to this modification are useful in ChIP assays (see below) to determine whether a gene is active or not. Several histone kinases have been identified (Rsk2, Jil-1, Msk1) and while not much is known about them, their activities are generally correlated with gene activation.

Ubiquitination

This type of modification occurs mostly on H2A, but has been observed on H2B and H3 tails. While not much is known about the functional consequences of this type of modification, it is known that ubiquitination is dependent upon on-going transcription - and thus it is associated with active genes, rather than protein degradation. Recently the TAF_{II}250 protein has been shown to be a ubiquitinating enzyme. TAF_{II}250 can ubiquitinate histone H1 *in vitro*. A mutation that kills ubiquitinating activity is down *in vivo*; the *Drosophila* transcription factor Dorsal is unable to activate certain target genes (Pham, A.D. and Sauer, F. 2000. *Science* 289:2357-2360.)

Methylation

To be expanded in class (see ppt file). Methylation is most commonly linked to gene repression and DNA methylation and the spreading of heterochromatin. However, methylation of lysine9 on histone H3 correlates strongly with the gene inactivity or silencing, methylation of lysine4 on the same histone is highly enriched at the 5' ends of transcribed genes. Most people assume that H3K4 marks actively transcribed genes and H3K9 marks silenced genes. As background for one of the assigned papers, methylation of H3K36 occurs by action of a histone methyltransferase Set2 which happens to bind specifically to the phosphorylated tail of RNA polymerase II and travels with the elongating form. It has been somewhat of a curiosity that actively transcribed genes are methylated on residue K36 – the functional consequence or “code” this imparts unknown. The paper by Carrozza et al. (Carrozza et al. 2005. *Cell* 123: 581-592) shows that this mark or code is recognized by a chromodomain of a protein (Eaf3) component of a histone de-acetylating complex (RPD3S). Why would you want to recruit an HDAC activity to an actively transcribed gene? The data presented in the paper show that aberrant, cryptic transcription initiation occurs along the gene unless the chromatin is actively de-acetylated. Thus methylation of H3K36 is a safety mechanism to limit transcription initiation to proper sites in promoters.

Histone De-acetylation

Just as there are enzymes to attach acetyl groups to histone tails, so there are enzymes that strip them off. These are the **HDACs**. The first HDAC was identified in human extracts on the basis of its ability to bind the deacetylase inhibitor drug trapoxin. It was given the name HDAC1. Purification and cloning of the gene encoding this enzyme showed aa sequence homology to a known transcription regulator in yeast called Rpd3. This was the first link between HDAC activity and transcription. Like HATs, HDACs are typically part of multi-subunit complexes. The number of HDACs and complexes is not as extensive as that for HATs (eight in human systems), but for here, we will only mention one notable complex - the NuRD complex. NuRD contains HDAC1 and HDAC2 and also a SWI/SNF-like ATPase. Thus, deacetylation occurs in a ATP-dependent manner. NuRD is recruited to regions of DNA that are heavily methylated. Methylated-DNA binding proteins, such as MeCP2, can directly recruit these complexes. Methylated DNA is tightly associated with silencing of genes.

Histone De-methylation

Methylation of histones is very stable and for a while people assumed that it was a permanent mark, only erased from chromatin by histone/nucleosome replacement. Imagine the surprise and excitement when two years ago, the protein LSD1 was identified as de-methylating protein (Shi et al. 2004. *Cell* 119:941-953). LSD1 can be shown to de-methylate mono- and di- methylated lysines but not tri-methylated residues.

F. Chromatin Remodeling

Remodeling involves the breaking and reforming of histone•DNA contacts and allows nucleosome mobilization. The precise mechanism and mode of movement or sliding is unknown. Remodeling complexes are distinct from the HAT and HDAC complexes discussed above (with the exception of NuRD). As you would expect me to say, there are by now numerous complexes identified, but we will only focus on one: SWI/SNF.

SWI/SNF was the first remodeling complex to be identified and studied - it was found with genetic screens in yeast to find components involved in regulation of transcription. Again, all remodeling machines are multi-subunit enzymes, and they all contain at least one ATPase subunit. Indeed, the ability to promote nucleosome sliding is ATP-dependent. The other subunits in the complex appear to regulate specificity, efficiency and stability of the complex. Mammals have 2 distinct SWI/SNF-like complexes that share many of the same subunits but are distinguished by their ATPase subunits: BRF1 and BRM.

These complexes appear to regulate the interconversion between chromatin states. This means that although remodeling complexes can accelerate the conversion to open chromatin, there are examples that they can accelerate conversion to closed chromatin too. Indeed, genome-wide array analysis shows that knock-out of remodeling subunits has both positive and negative effects on transcription. One might think of these complexes as being "state" neutral and only facilitators to a specified state but this is a generalization. With that said, most of the remodeling complexes have been examined in relationship to open chromatin and gene activation.

SWI/SNF is recruited to promoters by DNA-binding activators, including nuclear hormone receptors and it has been shown to remodel nucleosomes upon recruitment. SWI/SNF interacts with other transcription factors such as: GCN4, SWI5, Gal4-VP16, GAL4-AH. Also, SWI/SNF has been co-purified with holoRNAPolIII by several groups suggesting that it can at times associate with basal machinery, but this remains to be confirmed rigorously.

G. So who's on first?

So how are these two types of chromatin modification machines recruited to genes, and who comes in first? Do these machines stay around the promoter, or does the entire gene open up? Because we are relatively new to these machines and their modes of activity, only a couple of genes have been studied. What is clear is that generally, both types of activity are needed for gene activation. However, their order of recruitment differs somewhat. In yeast, at the mating type loci, recruitment of SWI/SNF by the DNA binding protein Swi5 is required before the GCN5/SAGA complex can be recruited. At other genes, only GCN5/SAGA is required, unless it is during mitosis and then the SWI/SNF machine is required. So, opening up chromatin and activating genes may be influenced by cell cycle and the overall state of chromatin in the first place. This topic, ie. what complexes are needed and when - is a new area of research in polIII transcription.

H. Elongation

Estimates of PolIII speed in vivo vary from 25-100 nt/second, which isn't very fast if you compare it to DNA polymerase and the replication complex. Studies of RNA polymerase movement on a single nucleosome template show that its speed is varied, slowing down considerably halfway through DNA wrapped around a nucleosome. Clearly, elongation would be served if there were some traveling modification activity.

Several different proteins and complexes have been shown to facilitate efficient transcription on nucleosomal templates in vitro. They are:

- SWI/SNF
- HMG14 - a chromatin associated binding protein - has a modest effect
- FACT - a multi-subunit activity that appears to strip H2A/H2B dimers from nucleosomes - at least transiently. Indeed, old studies of actively transcribed chromatin showed that the nucleosomes in these regions were down for H2A/H2B content.
- Spt4, 5, 6 proteins in yeast. These proteins have been shown to facilitate elongation in vitro, although their mode of action isn't known. Genetic disruption of these genes exhibit phenotypes consistent with defects in transcription elongation
- Elongator - a three subunit complex that specifically binds to phospho-CTD
 - one of the subunits of Elongator is a HAT named: E1p3
 - this HAT travels with the elongating RNA polII via elongator and can acetylate all four histones in the nucleosome

Acetylation of nucleosomes per se, is not likely to allow faster transcription. Nucleosome composition and DNA contacts are not likely to be altered enough. Rather, specific modifications may recruit proteins binding specifically to these modifications and help maintain an "open" structure along the entire length of the gene. Perhaps these modifications help recruit HMG14, or FACT or one of the Spt proteins

I. The structural dynamics of an actively transcribed gene

If transcriptional elongation through a gene brings in its wake complexes that acetylate nucleosomes leading to deposition of new proteins that maintain an open chromatin structure, then the following questions are important to consider:

- a. is active transcription needed to maintain an open structure
 - b. is there a "pioneer" polymerase that has to do the hard to work of the first round of transcription? Does it look different than polymerases that clear a re-initiating promoter?
- a. modification of chromatin by HATs that track with the polymerase must lead to de-repression or open chromatin of the whole transcription unit. If so, then one can imagine how genes might be turned off. Competition with HDACs in the nucleus for access to, and modification of, histone tails might require constant transcription. This is certainly supported by early experiments that showed active transcription and histone acetylation are required for open chromatin structures (operationally defined as sensitive to DNAaseI). Danny Reinberg proposes that the state of histone tail acetylation is a "dynamic equilibrium determined by the activities of HATs bound to elongating RNAP II and HDACs". In this model a decrease in gene transcription (usually specified by regulatory events at the promoter) would shift this dynamic to favor HDAC stripping of acetyl groups. This would be predicted to lead to rapid closing of the chromatin structure and overall gene repression.
- b. "pioneer polymerases"
- so what do transcription factors do? Do they promote changes in chromatin structure only over the promoter? Or do they nucleate a spreading change in chromatin that covers the entire gene? These two models differ in that in the first model, the first, initiating polymerase has quite a job to do. It must transcribe through compacted chromatin. It would help itself greatly by carrying along chromatin modification and remodeling complexes. There is evidence for this. In the second model, the initiating polymerase has an easier job. The paper assigned to you by Hahn and his group implies that the re-

initiating polymerase may indeed be different (remember? Re-initiating polymerases don't necessarily need to come in with mediator), but it is too early to tell which is the right model.