

development in the sea urchin have delineated a regulatory network that has significant predictive power<sup>16</sup> (Fig. 2). Finally, systems approaches to metabolism in an archaeal halobacterium (an organism thriving in up to five-molar salt solutions, such as the Dead Sea) have revealed new insights into the inter-relationships among several modules controlling energy production in the cell<sup>18</sup>.

The study of cellular and organismal biology using the systems approach is at its very beginning. It will require integrated teams of scientists from across disciplines — biologists, chemists, computer scientists, engineers, mathematicians and physicists. New methods for acquiring and analysing high-throughput biological data are needed. A powerful computational infrastructure must be leveraged to generate more effective approaches to the capture, storage, analysis, integration, graphical display and mathematic formulation of biological complexity. New technologies must be integrated with each other. Finally, hypothesis-driven and discovery science must be integrated. In short, both new science and technology must emerge for the systems biology approach to realize its promise. A cultural shift in the biological sciences is needed, and the education and training of the next generation of biologists will require significant reform.

Gordon Moore, the founder of Intel, predicted that the number of transistors that could be placed on a computer chip would double every 18 months. It has for more than 30 years. This exponential growth has been a driver for the explosive growth of information technology. Likewise, the amount of DNA sequence information available to the scientific community is following a similar, perhaps even steeper, exponential increase. The critical issue is how sequence information can be converted into knowledge of the organism and how biology will change as a result. We believe that a systems approach to biology is the key. It is clear, however, that this approach poses significant challenges, both scientific and cultural<sup>19</sup>. The discovery of DNA structure started us on this journey, the end of which will be the grand unification of the biological sciences in the emerging, information-based view of biology. □

doi:10.1038/nature01410

1. Watson, J. D. & Crick, F. H. C. A structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).
2. Brenner, S., Jacob, F. & Meselson, M. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* **190**, 576–581 (1961).
3. Crick, F. H. C., Barnett, L., Brenner, S. & Watts-Tobin, R. J. General nature of the genetic code for proteins. *Nature* **192**, 1227–1232 (1961).
4. Nirenberg, M. W. & Matthaei, J. H. The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polynucleotides. *Proc. Natl Acad. Sci. USA* **47**, 1588–1602 (1961).
5. Saiki, R. K. *et al.* Enzymatic amplification of  $\beta$ -globin sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**, 1350–1354 (1985).
6. Maxam, A. M. & Gilbert, W. A new method of sequencing DNA. *Proc. Natl Acad. Sci. USA* **74**, 560–564 (1977).
7. Sanger, F. & Coulson, A. R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* **94**, 444–448 (1975).
8. Sanger, F. *et al.* Nucleotide sequence of bacteriophage  $\phi$ X174. *Nature* **265**, 678–695 (1977).
9. Hunkapiller, M. W. & Hood, L. New protein sequenator with increased sensitivity. *Science* **207**, 523–525 (1980).
10. Horvath, S. J., Firca, J. R., Hunkapiller, T., Hunkapiller M. W. & Hood, L. An automated DNA synthesizer employing deoxynucleoside 3' phosphoramidites. *Methods Enzymol.* **154**, 314–326 (1987).
11. Kent, S. B. H., Hood, L. E., Beilan, H., Meister S. & Geiser, T. High yield chemical synthesis of biologically active peptides on an automated peptide synthesizer of novel design. *Peptides* **5**, 185–188 (1984).
12. Smith, L. M. *et al.* Fluorescence detection in automated DNA sequence analysis. *Nature* **321**, 674–679 (1986).
13. Collins, F. & Galas, D. J. A new five-year plan for the US Human Genome Project. *Science* **262**, 43–46 (1993).
14. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
15. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
16. Davidson, E. H. *et al.* A genomic regulatory network for development. *Science* **295**, 1669–1678 (2002).
17. Ideker, T. *et al.* Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929–933 (2001).
18. Baliga, N. S. *et al.* Coordinate regulators of energy transduction modules in the *Halobacterium* sp. analyzed by a global systems approach. *Proc. Natl Acad. Sci. USA* **99**, 14913–14918 (2002).
19. Aderem A. & Hood, L. Immunology in the post-genomic era. *Nature Immunol.* **2**, 1–3 (2001).
20. Dennis, C. & Gallagher, R. (eds) *The Human Genome* (Palgrave, Basingstoke, 2001).

# Controlling the double helix

Gary Felsenfeld\* & Mark Groudine†

\*Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Building 5, Room 212, Bethesda, Maryland 20892-0540, USA (e-mail: gary.felsenfeld@nih.gov)

†Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, Washington 98109, and Department of Radiation Oncology, University of Washington School of Medicine, Seattle, Washington 98195, USA (e-mail: markg@fhcr.org)

**Chromatin is the complex of DNA and proteins in which the genetic material is packaged inside the cells of organisms with nuclei. Chromatin structure is dynamic and exerts profound control over gene expression and other fundamental cellular processes. Changes in its structure can be inherited by the next generation, independent of the DNA sequence itself.**

**G**enes were first shown to be made of DNA only nine years before the structure of DNA was discovered (ref. 1; and see article in this issue by McCarty, page 406). Although revolutionary, the idea that genetic information was protein-free ultimately proved too simple. DNA in organisms with nuclei is in fact coated with at least an equal mass of protein, forming a complex called chromatin, which controls gene activity and the inheritance of traits.

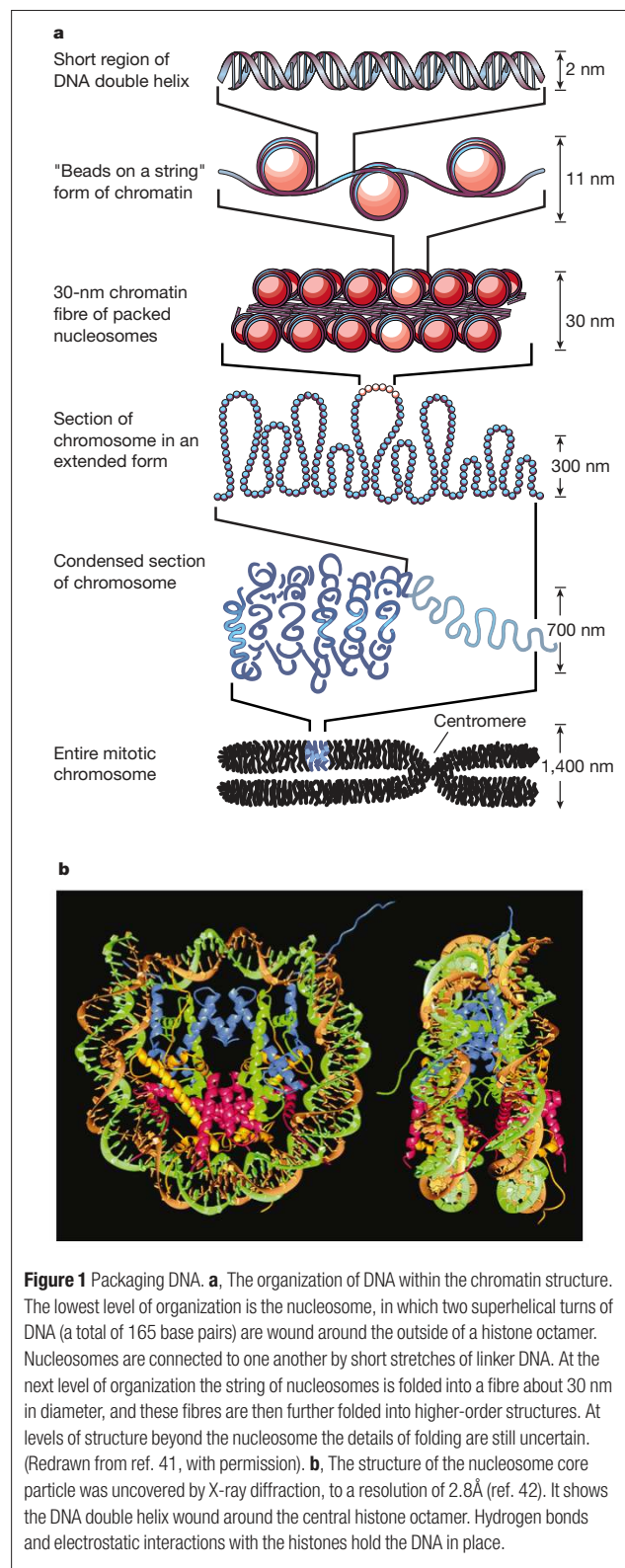
'Higher' organisms, such as yeast and humans, are eukaryotes; that is, they package their DNA inside cells in a separate compartment called the nucleus. In dividing cells, the chromatin complex of DNA and protein can be seen as individual compact chromosomes; in non-dividing cells, chromatin appears to be distributed throughout the nucleus and organized into 'condensed' regions (heterochromatin) and more open 'euchromatin' (see article in this issue by Ball, page 421). In contrast, prokaryotes, such as bacteria, lack nuclei.

## The evolution of chromatin

The principal protein components of chromatin are proteins called histones (Fig. 1). Core histones are among the most highly conserved eukaryotic proteins known, suggesting that the fundamental structure of chromatin evolved in a common ancestor of eukaryotes. Moreover, histone equivalents and a simplified chromatin structure have also been found in single-cell organisms from the kingdom Archaeobacteria<sup>2,3</sup>.

Because there is more DNA in a eukaryote than in a prokaryote, it was naturally first assumed that the purpose of histones was to compress the DNA to fit within the nucleus. But subsequent research has dramatically revised the view that histones emerged as an afterthought, forced on eukaryotic DNA as a consequence of large genome size and the constraints of the nucleus.

It was known that different genes are active in different tissues, and the distinction of heterochromatin and euchromatin suggested that differences in chromatin structure were associated with differences in gene expression. This led to the early supposition that the histones were also repressor proteins designed to shut off unwanted expression. The available evidence, although rudimentary, does indeed suggest that archaeal histones are not merely packaging factors, but function to regulate gene expression<sup>2-5</sup>. They



**Figure 1** Packaging DNA. **a**, The organization of DNA within the chromatin structure. The lowest level of organization is the nucleosome, in which two superhelical turns of DNA (a total of 165 base pairs) are wound around the outside of a histone octamer. Nucleosomes are connected to one another by short stretches of linker DNA. At the next level of organization the string of nucleosomes is folded into a fibre about 30 nm in diameter, and these fibres are then further folded into higher-order structures. At levels of structure beyond the nucleosome the details of folding are still uncertain. (Redrawn from ref. 41, with permission). **b**, The structure of the nucleosome core particle was uncovered by X-ray diffraction, to a resolution of 2.8 Å (ref. 42). It shows the DNA double helix wound around the central histone octamer. Hydrogen bonds and electrostatic interactions with the histones hold the DNA in place.

may facilitate gene activation, by promoting specific structural interactions between distal sequences, or repression, by occluding binding sites for transcriptional activators.

We suggest that the function of archaeal histones reflects their ancestral function, and therefore that chromatin evolved originally as an important mechanism for regulating gene expression. Its use in

packaging DNA was an ancillary benefit that was recruited for the more complex nucleosome structure that subsequently evolved in the ancestors of modern eukaryotes, which had expanded genome sizes. Although their compactness might seem to suggest inertness, chromatin structures are in fact a centre for a range of biochemical activities that are vital to the control of gene expression, as well as DNA replication and repair.

### Packaging DNA into chromatin

The fundamental subunit of chromatin is the nucleosome, which consists of approximately 165 base pairs (bp) of DNA wrapped in two superhelical turns around an octamer of core histones (two each of histones H2A, H2B, H3 and H4). This results in a five- to tenfold compaction of DNA<sup>6</sup>. The DNA wound around the surface of the histone octamer (Fig. 1) is partially accessible to regulatory proteins, but could become more available if the nucleosome could be moved out of the way, or if the DNA partly unwound from the octamer. The histone 'tails' (the amino-terminal ends of the histone protein chains) are also accessible, and enzymes can chemically modify these tails to promote nucleosome movement and unwinding, with profound local effects on the chromatin complex.

Each nucleosome is connected to its neighbours by a short segment of linker DNA (~10–80 bp in length) and this polynucleosome string is folded into a compact fibre with a diameter of ~30 nm, producing a net compaction of roughly 50-fold. The 30-nm fibre is stabilized by the binding of a fifth histone, H1, to each nucleosome and to its adjacent linker. There is still considerable debate about the finer points of nucleosome packing within the chromatin fibre, and even less is known about the way in which these fibres are further packed within the nucleus to form the highest-order structures.

### Chromatin regulates gene expression

Regulatory signals entering the nucleus encounter chromatin, not DNA, and the rate-limiting biochemical response that leads to activation of gene expression in most cases involves alterations in chromatin structure. How are such alterations achieved?

The most compact form of chromatin is inaccessible and therefore provides a poor template for biochemical reactions such as transcription, in which the DNA duplex must serve as a template for RNA polymerase. Nucleosomes associated with active genes were shown to be more accessible to enzymes that attack DNA than those associated with inactive genes<sup>7</sup>, which is consistent with the idea that activation of gene expression should involve selective disruption of the folded structure.

Clues as to how chromatin is unpacked came from the discovery that components of chromatin are subject to a wide range of modifications that are correlated with gene activity. Such modifications probably occur at every level of organization, but most attention has focused on the nucleosome itself. There are three general ways in which chromatin structure can be altered. First, nucleosome remodelling can be induced by complexes designed specifically for the task<sup>8</sup>; this typically requires that energy be expended by hydrolysis of ATP. Second, covalent modification of histones can occur within the nucleosome<sup>9</sup>. Third, histone variants may replace one or more of the core histones<sup>10–12</sup>.

Some modifications affect nucleosome structure or lability directly, whereas others introduce chemical groups that are recognized by additional regulatory or structural proteins. Still others may be involved in disruption of higher-order structure. In some cases, the packaging of particular genes in chromatin is required for their expression<sup>13</sup>. Thus, chromatin can be involved in both activation and repression of gene expression.

### Chromatin remodelling

Transcription factors regulate expression by binding to specific DNA control sequences in the neighbourhood of a gene. Although some DNA sequences are accessible either as an outward-facing segment on the nucleosome surface, or in linkers between nucleosomes, most

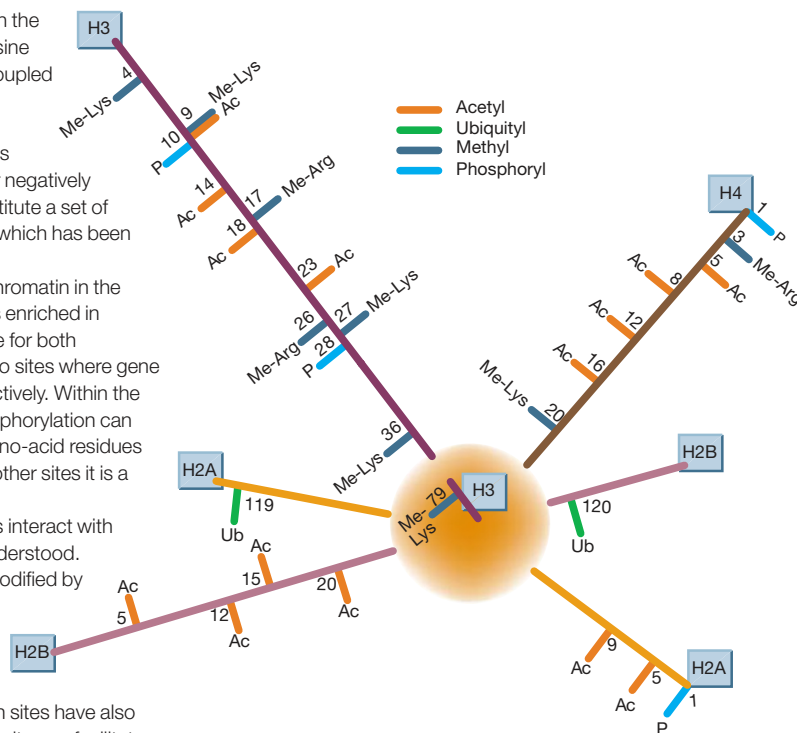
Box 1

**Histone modifications**

Many amino acids of histones, particularly those in the 'tails', are chemically modified<sup>47</sup>. These include lysine residues that may be acetylated, methylated or coupled to ubiquitin (a large polypeptide chain); arginine residues that may be methylated; and serine residues that are phosphorylated. All modifications can affect one another, and many are positively or negatively correlated with each other. Collectively, they constitute a set of markers of the local state of the genetic material, which has been called the 'histone code'<sup>48</sup>.

Histone modification is a dynamic process. Chromatin in the neighbourhood of transcriptionally active genes is enriched in acetylated histones, and the enzymes responsible for both acetylation and deacetylation are often recruited to sites where gene expression is to be activated or repressed, respectively. Within the nucleus, local states of both acetylation and phosphorylation can change rapidly. Methylation at certain histone amino-acid residues may also be important for activation, whereas at other sites it is a signal for inactivation.

Many (perhaps all) of the histone modifications interact with each other in ways that are still not completely understood. For example, in mammals, histone H2B can be modified by ubiquitin at Lys 120 (123 in yeast), and this modification is necessary for methylation at Lys 4 and Lys 79 of histone H3, reactions that are controlled by two different methylating enzymes. Influences between nearby modification sites have also been observed, such that phosphorylation at one site can facilitate acetylation at another, methylation and phosphorylation at adjacent sites may interfere with one another, and methylation and acetylation cannot occur simultaneously on the same lysine residue.



**Box 1 Figure** Histone modifications. Each modification is colour coded as indicated and the position of the modified amino acid labelled<sup>47,49-51</sup>.

are buried inside the nucleosome. Regulatory factors must therefore seek out their specific DNA-binding sites and gain access to them. They are aided by chromatin-remodelling complexes that continually shuffle the positions of individual nucleosomes so that sites are randomly exposed for a fraction of time<sup>8,14</sup>.

A number of chromatin-remodelling complexes mobilize nucleosomes, causing the histone octamers to move short distances along the DNA<sup>8</sup>. Each complex carries a protein with ATPase activity, which provides the necessary energy. Many of these complexes are members of the so-called SWI/SNF family, which includes SWI/SNF in budding yeast and human, RSC in yeast, and Brahma in *Drosophila*. They have similar helicase-motif subunits, but varying co-factors within the complex. Another SWI/SNF subfamily is based on the helicase-domain protein ISWI, which combines with other proteins to form the complexes NURE, CHRAC and ACF in *Drosophila*, and RSF in humans. A third subfamily is based on the helicase motif protein Mi-2.

Remodelling complexes differ in the mechanisms by which they disrupt nucleosome structure, and they are associated with co-factors that allow them to interact selectively with other regulatory proteins that bind to specific DNA sequences. For example, only certain classes of transcription factors interact with the mammalian SWI/SNF remodelling complex. Thus remodelling complexes can be selective in the genes they modify, and transcription factors recruit these complexes as tools to gain access to chromatin.

**Histone modification**

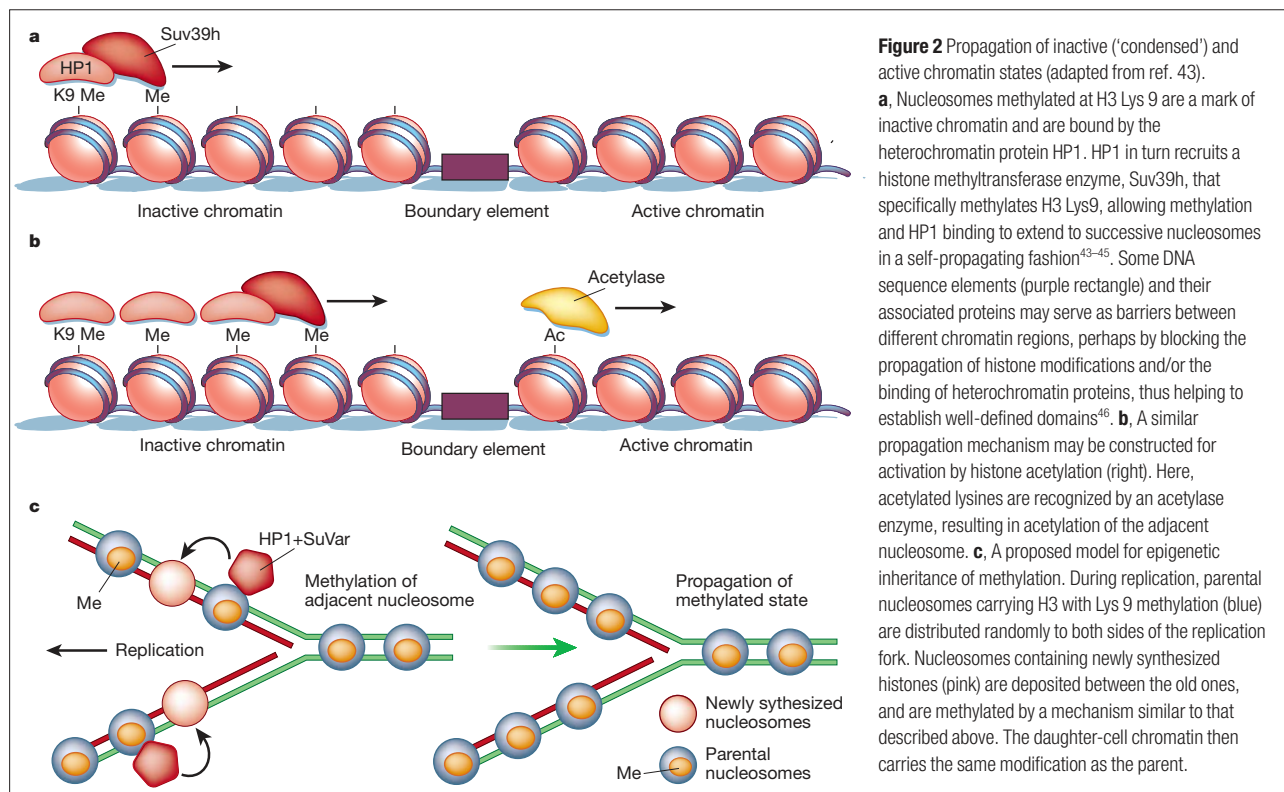
Nucleosomes are not passive participants in this recognition process. They can accommodate chemical modifications — either on histone

'tails' that extend from the nucleosome surface, or within the body of the octamer — that serve as signals for the binding of specific proteins. A large number of modifications are already known, such as acetylation of amino acids in the histone tails, and new ones are being identified at a bewildering rate (Box 1). Many modifications are associated with distinct patterns of gene expression, DNA repair or replication, and it is likely that most or all modifications will ultimately be found to have distinct phenotypes.

In addition to histone modifications, nucleosomes can have core histones substituted by a variant, with functional consequences. Histone H2AZ, which is associated with reduced nucleosome stability, replaces H2A non-randomly at specific sites in the genome. Histone H2AX, which is distributed throughout the genome, is a target of phosphorylation accompanying repair of DNA breakage<sup>11</sup>, and also seems to be involved in the *V(D)J* recombination events that lead to the assembly of immunoglobulin and T-cell-receptor genes. A histone H3 variant, H3.3, can be incorporated into chromatin in non-dividing cells, and seems to be associated with transcriptionally active genes<sup>10</sup>. Each of these histone substitutions is likely to be targeted by, and associated with, the binding of other proteins involved in gene activation; thus these proteins can be considered central to the formation of localized chromatin structures that are specific for gene activation or accessibility.

**Interdependence of histone modifications**

An interplay exists between histone modification and chromatin remodelling. For example, expression of a gene may require disruption of nucleosomes positioned at the promoter by a chromatin-remodelling complex before an enzyme required for histone



**Figure 2** Propagation of inactive ('condensed') and active chromatin states (adapted from ref. 43).

**a**, Nucleosomes methylated at H3 Lys 9 are a mark of inactive chromatin and are bound by the heterochromatin protein HP1. HP1 in turn recruits a histone methyltransferase enzyme, Suv39h, that specifically methylates H3 Lys9, allowing methylation and HP1 binding to extend to successive nucleosomes in a self-propagating fashion<sup>43–45</sup>. Some DNA sequence elements (purple rectangle) and their associated proteins may serve as barriers between different chromatin regions, perhaps by blocking the propagation of histone modifications and/or the binding of heterochromatin proteins, thus helping to establish well-defined domains<sup>46</sup>. **b**, A similar propagation mechanism may be constructed for activation by histone acetylation (right). Here, acetylated lysines are recognized by an acetylase enzyme, resulting in acetylation of the adjacent nucleosome. **c**, A proposed model for epigenetic inheritance of methylation. During replication, parental nucleosomes carrying H3 with Lys 9 methylation (blue) are distributed randomly to both sides of the replication fork. Nucleosomes containing newly synthesized histones (pink) are deposited between the old ones, and are methylated by a mechanism similar to that described above. The daughter-cell chromatin then carries the same modification as the parent.

acetylation can be recruited<sup>15</sup>. In contrast, expression of a different gene may require that histone-acetylating enzymes and even RNA polymerase bind to the promoter prior to recruitment of the chromatin-remodelling complex<sup>16</sup>. There is no common series of steps that underlies all or even most processes of gene activation. For any given gene, however, the order of recruitment of chromatin-modifying factors may be crucial for the appropriate timing of expression.

Aside from activating gene expression, histone modifications and chromatin remodelling can also silence genes. Specific histone modifications and chromatin-remodelling complexes, such as the NuRD complex, have been implicated in silencing at some loci<sup>8</sup>. Even SWI/SNF complexes, which are strongly correlated with gene activation, also seem to silence a number of genes.

### Specialized chromatin structures

Some regions of the genome are packaged in chromatin with distinct structural features. Three of the most studied such regions are centromeres (important for chromosomal organization during mitosis), telomeres (at the ends of chromosomes) and the inactive X chromosome in mammals. In each case, specific chromosomal structures are defined both by histones modified or substituted in specific patterns, and by the association of additional non-histone proteins or even by regulatory RNA molecules, which increasingly are implicated in chromatin organization<sup>17–19</sup>.

Inactive X chromosomes in mammals are enriched for the histone variant macroH2A<sup>20</sup>, which is almost three times as large as H2A itself. At vertebrate centromeres, one of the core histones, H3, is replaced by a variant, CENP-A; a similar replacement occurs in centromeres of the fruitfly *Drosophila*, indicating that this is an ancient evolutionary adaptation at centromeres. CENP-A in turn forms a complex with the centromere proteins CENP-B and -C, which mediates the formation of phased arrays of CENP-A-containing nucleosomes. In turn, additional proteins are recruited during cell division to enable the orderly separation of the two chromatids that make up each chromosome. After DNA replication, the sister chro-

matids are held together initially by a multisubunit complex called cohesin, while a second complex, condensin, helps to compact the chromosomes<sup>21</sup>. These complexes recognize distinct centromere structures, and a specialized nucleosome-remodelling complex associates with cohesin to help it gain access to the chromosomes<sup>22</sup>.

In the budding yeast *Saccharomyces cerevisiae*, gene silencing at the ends of chromosomes is mediated by a complex that assembles at telomeres. The complex is stabilized by the binding of the protein RAP1 to the telomere repeat sequences. Additional components, including the silent information regulator (SIR) proteins, then bind inward from the telomere ends, partly through interactions with local nucleosomes<sup>23</sup>. One of the SIR proteins is a histone deacetylase and is thought to repress gene expression at this site. Some components of these unique complexes are evolutionarily conserved, suggesting that these unusual chromatin structures may be found in organisms other than yeast.

The silencing of genes in the vicinity of centromeres in the fission yeast *Schizosaccharomyces pombe* has been shown recently<sup>17–19</sup> to depend on a set of RNA-processing enzymes involved in RNA interference, a process by which double-stranded RNA directs sequence-specific degradation of messenger RNA. One of these enzymes, Dicer, generates RNA fragments about 23 nucleotides long from transcripts of centromeric regions, which then seem in some way to be targeted back to the centromere to initiate the histone-dependent silencing mechanism. Moreover, non-coding RNA transcripts have been identified on the inactive X chromosome and elsewhere in the genome, and may have related roles at those loci<sup>24</sup>.

### Epigenetic inheritance

An epigenetic trait is one that is transmitted independently of the DNA sequence itself. This can occur at the level of cell division — for example, daughter cells may inherit a pattern of gene expression from parental cells (so-called cellular memory) — or at the generational level, when an offspring inherits a trait from its parents.

The classic example of epigenetic inheritance is the phenomenon of imprinting, in which the expression status of a gene depends upon

the parent from which it is derived. In mammals, for example, the *Igf2* gene (encoding insulin-like growth factor-2) is expressed only from the paternal copy of the gene, whereas the *H19* gene is expressed solely from the maternal allele. The mechanism by which this pattern of inheritance is accomplished involves (in part) DNA methylation on the paternal allele. This causes dissociation of a chromatin protein known as CTCF, which normally blocks a downstream enhancer; consequently, the enhancer is then free to activate *Igf2* expression<sup>25,26</sup>.

The methylation state of an allele is linked inextricably with patterns of histone modification<sup>27</sup>. Methylated CpG (guanine–cytosine) dinucleotide sites near a gene recruit specific DNA-binding proteins, which in turn recruit histone deacetylases, resulting in loss of histone acetylation and silencing of gene expression. But if histone deacetylation occurs first, it is possible to replace the acetyl group at histone H3 lysine 9 (Lys 9) with one to three methyl groups. It has been shown in turn in the fungus *Neurospora* that the ability to methylate histone Lys 9 is essential for DNA methylation<sup>28</sup>, suggesting that local methylation at Lys 9 may provide a signal for methylation of the underlying DNA. Furthermore, in a different reaction pathway, maintenance of histone acetylation at promoters can lead to inhibition of DNA methylation<sup>29</sup>.

Epigenetic inheritance involves the maintenance of patterns of histone modification and/or of association of chromosomal proteins correlated with specific expression states. The same mechanisms for propagating permissive or repressive chromatin structure could preserve the pattern of histone modification during replication, when old nucleosomes are distributed randomly on both sides of the fork, with the newly synthesized histones interspersed (Fig. 2).

The maintenance of repressed or activated transcription states represents an efficient mechanism for progressive cellular differentiation<sup>30</sup>. In such a model, fundamental decisions regarding the turning on or off of genes or groups of genes need to be made only once. This principle is perhaps most clearly illustrated by the example of Polycomb-group (PcG)-mediated gene repression in *Drosophila*<sup>31</sup>. At a specific time during development, a complex of proteins, encoded by a collective of PcG genes, binds to sequences within some genes, but only in cells where the genes are silent. At subsequent stages of development, the repressed state is maintained by the PcG complex in the absence of the original negative signals. Activated expression states can be similarly maintained, again in the absence of the original transcriptional activators, by a complex of proteins encoded by genes collectively termed the trithorax group<sup>31</sup>. In both cases, the maintenance of gene-expression patterns is associated with specific histone modification and chromatin-remodelling activities<sup>32–34</sup>.

### Chromatin and nuclear self-organization

Although bacteria lack a true nucleus, a specific region of the cell, called the nucleoid, contains the chromosome, which in turn is organized into supercoiled domains or loops emanating from central nodes. The organization of the *Escherichia coli* genome into such domains is necessary to allow it to fit within the confines of the cell<sup>2</sup>. Extensions of the chromosome into the cytoplasm correlate with regions that are transcriptionally active. Upon inhibition of transcription, these extensions recede to the nucleoid to give it a more even, spherical shape. The localization of genomic sequences within a bacterial cell is thus determined by their association with the transcriptional/translational apparatus.

The organization of the genome in eukaryotic nuclei, while necessarily more complex than in bacteria, seems to follow the same model as *E. coli*. Individual chromosomes largely occupy distinct ‘territories’ within the nucleus. Within these territories, actively transcribed genes are on surfaces of channels within subchromosomal domains<sup>35</sup> where soluble transcription factors are presumably more likely to gain access to them.

There is, however, more to the story. The eukaryotic nucleus has distinct subcompartments within which specific nuclear proteins are enriched. For example, the nucleolus, where high-level transcription

of ribosomal genes occurs, and splicing-factor compartments accumulate high local concentrations of certain proteins. In some cases there are attachment sites within the nucleus for the proteins. As a rudimentary example, one or more of the proteins associated with yeast telomeres is able to tether the telomeres in clusters to the nuclear periphery<sup>36</sup>. This clustering creates a high local concentration of binding sites for the SIR silencing proteins, which in turn results in a high local concentration of these proteins, and a high occupancy of even relatively weak binding sites. The effect is to increase the extent of telomeric silencing — SIR-dependent gene silencing can be accomplished just by artificially tethering a gene to the nuclear periphery<sup>37</sup>.

What organizes the formation of nuclear subdomains? Although there is evidence for a proteinaceous nuclear matrix<sup>38</sup>, the example provided by yeast telomeres suggests that the chromatin fibre itself may be the organizer. Many, and probably most, chromatin-binding proteins are in continuous flux between association with chromatin and the nucleoplasm<sup>39,40</sup>. Even such fundamental chromatin proteins as histone H1 have been found to bind for periods of only a few seconds, interspersed with periods of free diffusion. The notable exceptions to this rule are the core histones, the binding of which is much more stable — on the order of minutes for H2A/H2B, and hours for H3/H4. The on–off rates of proteins binding different regions of the genome may depend on the pattern of histone modifications, which in turn determines their relative enrichment in different regions of the nucleus. Thus, the genome as packaged with histones could determine the nature of nuclear subcompartments.

### Future challenges

Chromatin proteins and DNA are partners in the control of the activities of the genetic material within cells. The rate-limiting step in activating gene expression typically involves alterations of chromatin structure. The chromosome is an intricately folded nucleoprotein complex with many domains, in which local chromatin structure is devoted to maintaining genes in an active or silenced configuration, to accommodating DNA replication, chromosome pairing and segregation, and to maintaining telomeric integrity. Recent results suggest strongly that in all of these cases the primary indicators of such specialization are carried on the histones. Thus, the regulatory signals that determine local properties, as well as epigenetic transmission of those properties, are likely to be on histones.

The already large catalogue of histone modifications continues to grow rapidly. Although in most cases the loss of the modification (for example, by mutating the responsible enzyme) has a detectable effect on phenotype, the function of many modifications has not yet been determined. While this will be the focus of future research, it presents significant problems because a given modification will occur at many sites in the genome, and mutations could have widespread effects, both direct and indirect. A second significant challenge arises from the potential redundancy of the ‘histone code’: it is possible that either of two distinct modifications could specify a single structural and functional state, or that the two modifications are always linked to one another. Significant effort will be necessary to determine the complexity of this code, that is, the number of distinct states that can be specified.

The most important immediate problem is to identify the initiating step in establishing a local chromatin state, which may also correspond to an epigenetic state. Silencing at centromeres and perhaps elsewhere seems to be initiated by small RNA transcripts from within the region to be silenced, but formation of other kinds of structures might be triggered directly by a specific histone modification. In the longer term it will be necessary to relate the reactions at individual nucleosomes to higher-order chromatin structures; this will depend in part on the development of higher-resolution methods for determining those structures, and their organization within the nucleus.

At its simplest level, chromatin should be viewed as a single entity, carrying within it the combined genetic and epigenetic codes.

Ultimately our understanding of the dynamic states of chromatin throughout the genome will be integrated with a detailed knowledge of patterns of regulation of all genes. □

doi:10.1038/nature01411

- Avery, O. T., MacLeod, C. M. & McCarty, M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a desoxyribonucleic acid fraction isolated from *Pneumococcus* Type III. *J. Exp. Med.* **79**, 137–158 (1944).
- Sandman, K., Pereira, S. L. & Reeve, J. N. Diversity of prokaryotic chromosomal proteins and the origin of the nucleosome. *Cell Mol. Life Sci.* **54**, 1350–1364 (1998).
- Ouzounis, C. A. & Kyriakidis, N. C. Parallel origins of the nucleosome core and eukaryotic transcription from Archaea. *J. Mol. Evol.* **42**, 234–239 (1996).
- Bailey, K. A. & Reeve, J. N. DNA repeats and archaeal nucleosome positioning. *Res. Microbiol.* **150**, 701–709 (1999).
- Dinger, M. E., Baillie, G. J. & Musgrave, D. R. Growth phase-dependent expression and degradation of histones in the thermophilic archaeon *Thermococcus zilligii*. *Mol. Microbiol.* **36**, 876–885 (2000).
- Kornberg, R. D. Chromatin structure: a repeating unit of histones and DNA. *Science* **184**, 868–871 (1974).
- Weintraub, H. & Groudine, M. Chromosomal subunits in active genes have an altered conformation. *Science* **193**, 848–856 (1976).
- Becker, P. B. & Horz, W. ATP-dependent nucleosome remodeling. *Annu. Rev. Biochem.* **71**, 247–273 (2002).
- Zhang, Y. & Reinberg, D. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev.* **15**, 2343–2360 (2001).
- Ahmad, K. & Henikoff, S. Histone H3 variants specify modes of chromatin assembly. *Proc. Natl. Acad. Sci. USA* **10.1073/pnas.172403699** (2002).
- Redon, C. *et al.* Histone H2A variants H2AX and H2AZ. *Curr. Opin. Genet. Dev.* **12**, 162–169 (2002).
- Smith, M. M. Centromeres and variant histones: what, where, when and why? *Curr. Opin. Cell Biol.* **14**, 279–285 (2002).
- Wolffe, A. P. Nucleosome positioning and modification: chromatin structures that potentiate transcription. *Trends Biochem. Sci.* **19**, 240–244 (1994).
- Narlikar, G. J., Fan, H. Y. & Kingston, R. E. Cooperation between complexes that regulate chromatin structure and transcription. *Cell* **108**, 475–487 (2002).
- Cosma, M. P., Tanaka, T. & Nasmyth, K. Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle- and developmentally regulated promoter. *Cell* **97**, 299–311 (1999).
- Agalioti, T. *et al.* Ordered recruitment of chromatin modifying and general transcription factors to the IFN- $\beta$  promoter. *Cell* **103**, 667–678 (2000).
- Hall, I. M. *et al.* Establishment and maintenance of a heterochromatin domain. *Science* **297**, 2232–2237 (2002).
- Volpe, T. A. *et al.* Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**, 1833–1837 (2002).
- Allshire, R. RNAi and heterochromatin—a hushed-up affair. *Science* **297**, 1818–1819 (2002).
- Chadwick, B. P. & Willard, H. F. Cell cycle-dependent localization of macroH2A in chromatin of the inactive X chromosome. *J. Cell Biol.* **157**, 1113–1123 (2002).
- Nasmyth, K. Segregating sister genomes: the molecular biology of chromosome separation. *Science* **297**, 559–565 (2002).
- Hakimi, M. A. *et al.* A chromatin remodelling complex that loads cohesin onto human chromosomes. *Nature* **418**, 994–998 (2002).
- Grunstein, M. Molecular model for telomeric heterochromatin in yeast. *Curr. Opin. Cell Biol.* **9**, 383–387 (1997).
- Kelley, R. L. & Kuroda, M. I. Noncoding RNA genes in dosage compensation and imprinting. *Cell* **103**, 9–12 (2000).
- Bell, A. C. & Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* **405**, 482–485 (2000).
- Hark, A. T. *et al.* CTCF mediates methylation-sensitive enhancer-blocking activity at the *H19/Igf2* locus. *Nature* **405**, 486–489 (2000).
- Richards, E. J. & Elgin, S. C. Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. *Cell* **108**, 489–500 (2002).
- Jackson, J. P., Lindroth, A. M., Cao, X. & Jacobsen, S. E. Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* **416**, 556–560 (2002).
- Mutskov, V. J., Farrell, C. M., Wade, P. A., Wolffe, A. P. & Felsenfeld, G. The barrier function of an insulator couples high histone acetylation levels with specific protection of promoter DNA from methylation. *Genes Dev.* **16**, 1540–1554 (2002).
- Weintraub, H., Flint, S. J., Leffak, I. M., Groudine, M. & Grainger, R. M. The generation and propagation of variegated chromosome structures. *Cold Spring Harb. Symp. Quant. Biol.* **42**, 401–407 (1978).
- Francis, N. J. & Kingston, R. E. Mechanisms of transcriptional memory. *Nature Rev. Mol. Cell Biol.* **2**, 409–421 (2001).
- Cavalli, G. & Paro, R. Epigenetic inheritance of active chromatin after removal of the main transactivator. *Science* **286**, 955–958 (1999).
- Sewalt, R. G. *et al.* Selective interactions between vertebrate polycomb homologs and the SUV39H1 histone lysine methyltransferase suggest that histone H3-K9 methylation contributes to chromosomal targeting of Polycomb group proteins. *Mol. Cell Biol.* **22**, 5539–5553 (2002).
- Cao, R. *et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039–1043 (2002).
- Cremer, T. & Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature Rev. Genet.* **2**, 292–301 (2001).
- Laroche, T. *et al.* Mutation of yeast Ku genes disrupts the subnuclear organization of telomeres. *Curr. Biol.* **8**, 653–656 (1998).
- Andrulis, E. D., Neiman, A. M., Zappulla, D. C. & Sternglanz, R. Perinuclear localization of chromatin facilitates transcriptional silencing. *Nature* **394**, 592–595 (1998).
- Hart, C. M. & Laemmli, U. K. Facilitation of chromatin dynamics by SARs. *Curr. Opin. Genet. Dev.* **8**, 519–525 (1998).
- Misteli, T. Protein dynamics: implications for nuclear architecture and gene expression. *Science* **291**, 843–847 (2001).
- Hager, G. L., Elbi, C. & Becker, M. Protein dynamics in the nuclear compartment. *Curr. Opin. Genet. Dev.* **12**, 137–141 (2002).
- Alberts, B. *et al.* *Essential Cell Biology: An Introduction to the Molecular Biology of the Cell* (Garland, New York, 1998).
- Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
- Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120–124 (2001).
- Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116–120 (2001).
- Aagaard, L. *et al.* Functional mammalian homologues of the *Drosophila* PEV-modifier *Su(var)3-9* encode centromere-associated proteins which complex with the heterochromatin component M31. *EMBO J.* **18**, 1923–1938 (1999).
- West, A. G., Gaszner, M. & Felsenfeld, G. Insulators: many functions, many mechanisms. *Genes Dev.* **16**, 271–288 (2002).
- Goll, M. G. & Bestor, T. H. Histone modification and replacement in chromatin activation. *Genes Dev.* **16**, 1739–1742 (2002).
- Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).
- Wolffe, A. P. & Hayes, J. J. Chromatin disruption and modification. *Nucleic Acids Res.* **27**, 711–720 (1999).
- van Leeuwen, F., Gafken, P. R. & Gottschling, D. E. Dot1p modulates silencing in yeast by methylation of the nucleosome core. *Cell* **109**, 745–756 (2002).
- Ng, H. H. *et al.* Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association. *Genes Dev.* **16**, 1518–1527 (2002).

#### Acknowledgements

We thank the members of our laboratories, L. Hartwell and W. Bickmore for their advice and helpful comments. In particular, we are indebted to M. Bulger for his help in clarifying many of the issues addressed. We also acknowledge the seminal contributions to this field made by our late colleague and friend H. Weintraub, whose ideas continue to serve as a guide to our thinking.