such as N1 and nucleoplasmin instead for two reasons: first, to carry the histones from their site of synthesis on the ribosome, across the nuclear membrane to their site of assembly on the chromosome; and second, to pick out H3, H4 and H2A, H2B from all the different positively charged proteins in the cell, that might be able to bind to DNA non-specifically.

The discovery of these histone carriers was good progress, but then a stumbling block was reached. When the complexes of N1 with H3, H4, and nucleoplasmin with H2A, H2B were purified from the cell extract, it turned out that the histones would assemble onto DNA in the test-tube at a spacing of only 145 base-pairs, rather than at a spacing of 180 base-pairs as in the original extract from frog eggs. In other living cells, the spacing can vary from as low as 160 to as high as 260 base-pairs, depending on the proteins present in any nucleosome; but one never sees a very short spacing of 145 base-pairs in Nature.

Figure 7.8 shows some typical results for the assembly of histones onto DNA by an extract from living cells on the one hand, and by a pure system on the other. The procedure for determining the spacing of nucleosomes is to 'digest' the preparation by means of a DNA-cutting enzyme (like the one used by Hewish and Burgoyne), and then to measure the sizes of the fragments so obtained by using an electrophoretic gel according to a scheme which we shall describe in Chapter 9.

On the left-hand side of Fig. 7.8, one can see by 'reading' the gel that DNA may be assembled with histones from a frog-egg extract to make particles of size 180, 360, 540, 720, and 900 base-pairs; and so the size of the fundamental unit is 180 base-pairs. On the right-hand side of Fig. 7.8, DNA has been assembled with histones in a pure system (that is, without any cell extract), and it has made particles of size 145, 290, 435, 580, 725, and 870 base-pairs, after digestion with the enzyme; and so the size of the fundamental unit is 145 base-pairs. (The lowest band of size 145 base-pairs is obscured in this gel, but it can be seen in other experiments.)

One generally accepted explanation for the reduction in spacing from 180 to 145 base-pairs, on going from a cell extract to a pure system, is that we have lost certain important proteins when purifying the material. If these proteins normally associate with the histones as they assemble onto DNA in a cell, then in the cell extract there would be more proteins per particle than in the pure system, at least while they are being assembled; and hence each particle would extend for a greater length along the DNA. Which proteins might we have lost? The frog-egg extract is known not to contain any of the usual histone H1, so perhaps some other abundant protein (or proteins) might be important to the correct assembly of nucleosomes.

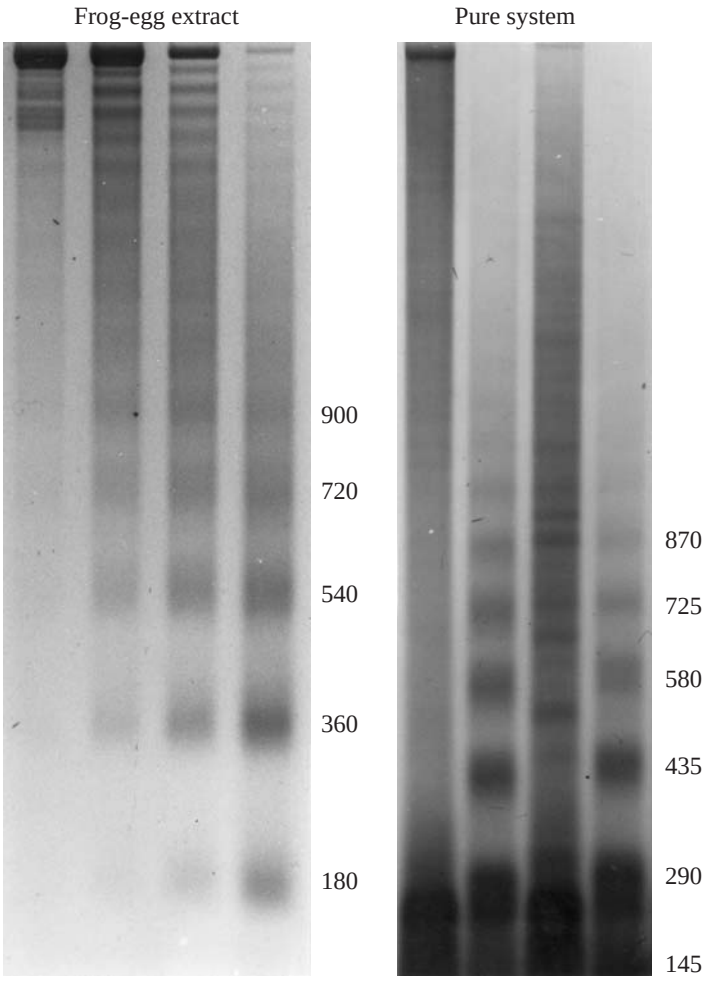

Figure 7.8 Analysis of nucleosome assembly by use of electrophoresis in gels, after digestion by enzymes. Histones are assembled onto DNA at a spacing of 145 base-pairs in a pure system, but at a more authentic spacing of 180 base-pairs in an extract from frog eggs. Left-hand gel by courtesy of David Tremethick.

The most likely candidate proteins for this task would be the high-mobility group proteins HMGA, HMGB and HMGN, which we mentioned earlier. These small proteins are fairly abundant in the cell nucleus, and they also bind to DNA with low specificity for the base sequence. Several experiments by David Tremethick have shown that either a moderately pure fraction of proteins taken from the frog-egg extract, and containing various HMG-like proteins, or else a completely pure preparation of phosphorylated proteins HMGN from human placenta, will increase the spacing of nucleosomes from 145 to 165 base-pairs. Other workers, for example, James Kadonaga and

colleagues, have found that HMGN will increase the spacing of nucleosomes in a fly-egg extract from 160 to 175 base-pairs per particle.

When histone H1 is added to a frog-egg or fly-egg extract, or to a partially-purified extract, the spacing of nucleosomes grows even larger: from about 180 to 220 base-pairs in a crude extract, or from 165 to 190 base-pairs in a partially-purified extract. Many independent experiments by various workers show that effect clearly. Most mature tissues in animals such as the fly, chicken, and human contain histone H1, and they also show spacings on the gel that correspond to a fundamental unit from 190 to 220 base-pairs long. Animals such as the sea urchin show a spacing of 260 base-pairs in some tissues; and perhaps those cells contain other special nucleosome-assembly proteins that have not yet been identified.

A tentative summary of these results is shown schematically in Fig. 7.9. It seems likely that the assembly of DNA into nucleosomes proceeds by at least three steps. First, the N1 protein (or some other negatively charged carrier) binds two copies each of histones H3, H4 and assembles them onto DNA. Second, the nucleoplasmin protein may bind H2A, H2B, and possibly an HMG-like protein, and add these proteins to DNA on either side of the already-assembled H3, H4 tetramer. The HMG-like protein may in some cases be related to HMGN. Without this HMG-like protein, during assembly there

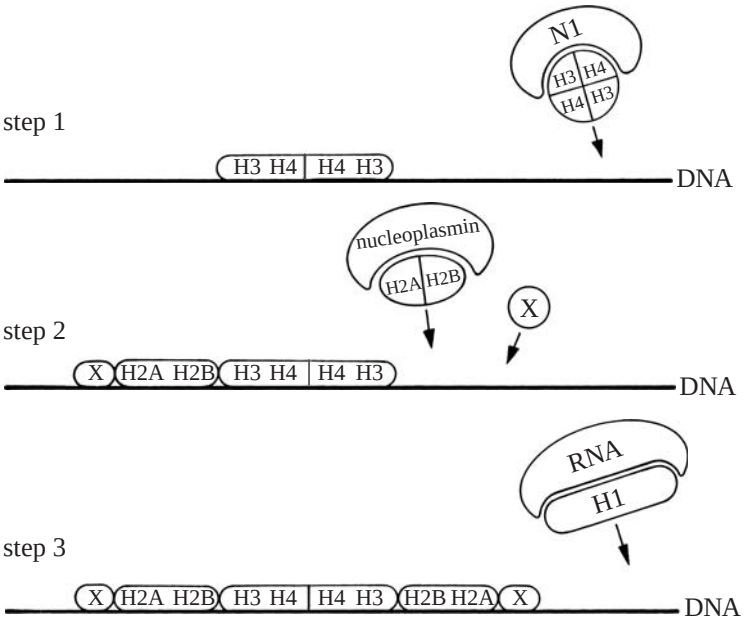

Figure 7.9 Three-step general scheme for the assembly of nucleosomes. The identity of the protein shown here as X is not yet known for certain; it is probably an HMG-like protein, or is perhaps related to HMGN, according to current work.

may be only eight proteins per particle, spanning 145 base-pairs of DNA; but with the HMG-like protein there are probably 10 proteins per particle during assembly, spanning 165 base-pairs. Finally, if histone H1 is present, it may add to the DNA between particles, to increase the spacing in the simplest case from 165 to 190 base-pairs, or in other cases from 180 to 220 base-pairs. The carrier for histone H1 in living cells is not known, but in the extract it seems to be some kind of RNA. No doubt the conjectured picture shown in Fig. 7.9 will be altered and refined in the future.

One important consequence of a stepwise assembly of DNA into nucleosomes, as in Fig. 7.9, is the following: each step in the procedure is reversible, so we can expect disassembly to proceed by the same series of steps but in a reverse direction. The relative rates of assembly *versus* disassembly within any step will therefore be of importance to specific biological processes, such as gene activation, that require the DNA to be at least partly unraveled from its folded state. We shall return to this subject later.

A large part of this book has been devoted to a study of DNA curvature. How does the curvature of DNA relate to its assembly into nucleosomes? In Figs 7.1 and 7.2 it was shown that DNA curves for almost two superhelical turns, each of about 80 base-pairs, around the histone proteins in any nucleosome. Thus, one might expect that *curved* or else very *flexible* DNA could assemble into nucleosomes more easily than DNA of mixed sequence, because it would take less energy for those special sequences to wrap into the required superhelical shape. Both of these factors are indeed important. Generally, even a small amount of curvature, as specified by the DNA base sequence (see Chapter 4), is sufficient to locate the path of DNA about the histone proteins. And also, in the absence of intrinsic curvature, the more flexible the DNA, the stronger is its affinity for histone proteins.

In addition, experiments using living cells have shown that the histone proteins often adopt highly ordered locations with respect to the DNA sequence in a cell nucleus, in a way that is influenced but not strictly determined by DNA flexibility. One can calculate the ease of curvature of any DNA sequence about a single set of histone proteins by Fourier algorithms similar to those described in Chapters 4 and 5. Yet it is difficult to calculate such things precisely over several successive sets of histone proteins, because successive nucleosomes may stick together, or fold into a 300 Å fiber, in a cooperative fashion that works against preferences in the DNA for curvature or flexibility. Recent experiments suggest however that the DNA sequence may specify the location of the first nucleosome in a long array, which then influences others nearby.

Several experiments have shown that the insertion of curved or else very flexible DNA near the start of a gene can actually repress the activity of that gene. In those cases it seems likely that the histone proteins bind to the DNA more tightly than usual, on account of curvature or flexibility; and so the specialized 'transcription factor' proteins which are used to initiate RNA synthesis cannot gain access to the DNA, in order to make the gene work. (The zinc-finger Zif268 protein described in Chapters 4 and 8 is one example of such a transcription factor.) Consistent with this idea, it is known that the insertion of non-curved or stiff DNA near the start of a gene (as for example long runs of sequence AA/TT or GG/CC) can actually activate that gene, by increasing the access of transcription factors to nearby DNA. Furthermore, when curved DNA is present on a large scale throughout an entire chromosome (as for example in certain birds and reptiles), then genes throughout the whole chromosome may become repressed, and the chromosome itself may remain so compactly folded and condensed, that it becomes visible at interphase.

Literally hundreds of experiments have been performed, where transcription factors and histone proteins (or HMG proteins) are added to the same piece of DNA, and the effects of one on the other measured. These studies are rather poor representations of what actually happens in the chromosome of a living cell, but they have turned up some interesting results. For example, certain transcription factors may bind to the same piece of DNA with widely different affinities, depending on which way the DNA curves about the histones, and hence which face of the DNA helix remains exposed. Those particular results seem to mimic well what happens in living cells, when transcription factors activate genes that are already wrapped in nucleosomes.

A broad conclusion that can be drawn from these studies is that the actions of DNA in a chromosome are very intricate and complex. If you try to study the workings of DNA solely in terms of abstractly defined genes and transcription factors, without paying attention to its three-dimensional structure, and to any proteins which are tightly associated with the DNA in a chromosome, then you will encounter many cellular phenomena which you will be unable to explain. In the past, cell-free studies on the mechanisms of transcription in higher organisms used free DNA, rather than DNA packaged into authentic pieces of a chromosome. But while testing the activities of such transcription factors, Kadonaga and colleagues found that the use of DNA bound to nucleosomal proteins, in a cell-free transcription assay, produced results that were more similar to those found in living cells, than by using as a template histone-free DNA.

A major precursor to the activation of any gene must be the unraveling of DNA from its tight folding about the histone proteins, at least transiently. Without such unravelling, RNA polymerase and other proteins required for 'reading' the DNA could never gain access to their required binding sites near the start of a gene. How, then, does the DNA become accessible to transcription factors and RNA polymerase, while at same time remaining associated with histones?

Our cells seem to have developed elaborate mechanisms for altering the local structure of chromatin, and for sliding or 'shuffling' nucleosomes along the DNA within active genes. In the test tube, nucleosomes possess an intrinsic but random ability to move along a piece of DNA. Yet in the living cell it seems that this ability is not used randomly, but instead is regulated actively. This implies that certain nucleosomes must carry some kind of 'flag' to indicate where such shuffling should take place.

The most likely mechanism for a 'flag' would be the chemical modification of histone proteins. Those histones H2A, H2B, H3 and H4 can be chemically modified in several different ways, the most important of which are acetylation, phosphorylation, and methylation. These three different modifications are performed by special enzymes known as *acetylases*, *kinases*, and *methylases* respectively. When a protein is acetylated, an acetyl group (CH_3CO) is added to one or more of its lysine amino acids, thereby removing the positive charge; see Fig. 7.10(a). When a protein is phosphorylated, a phosphate group (PO_3^-) is added, usually to one or more of its serine or threonine amino acids, thereby adding a negative charge; see Fig. 7.10(b). Finally, when a protein is methylated, either one, two or three methyl groups (CH_3) are added to its lysine amino acids,

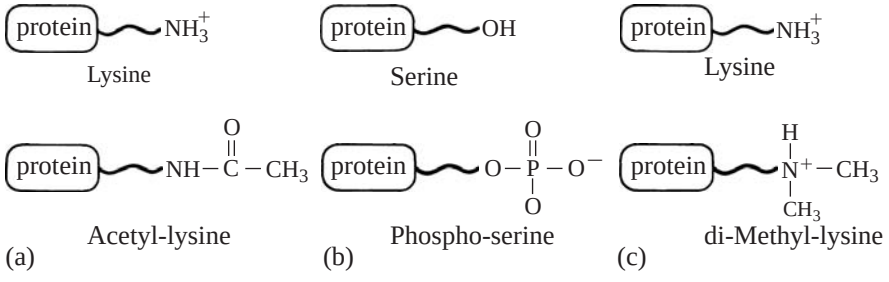

Figure 7.10 (a) Acetylation, (b) phosphorylation, and (c) methylation of typical amino acids in a protein. The first two modifications change the electric charge of the protein, in a way that may influence its structure and its ability to interact with DNA or with other proteins. In (c) two methyl groups have been attached; but a third might be added by replacing the H by a CH_3 group (or alternatively in the single-methyl form, just one of the three hydrogens could be replaced by a CH_3 group).

Table 7.1 The acetylation, phosphorylation or methylation of certain chromosomal proteins can affect their physical properties, and so influence their biological functions

Protein	Modification	Effect
H2A, H2B, H3, H4	Acetylation	Usually, easy sliding or shuffling on DNA
H3	Phosphorylation	Gene activation
H3	Methylation	Modify gene activity (see Chapter 11)
HMGB	Acetylation	Alters DNA binding
HMGN	Phosphorylation	Bind H2A, H2B more tightly
HMGA	Acetylation Phosphorylation Methylation	Not known
H1	Phosphorylation	Binds DNA less tightly
N1, nucleoplasmin	Phosphorylation	Bind histones more tightly

which increase their bulk but do not alter the electric charge; see Fig. 7.10(c). And arginine amino acids can also be modified by the addition of one methyl group.

The most important kinds of histone modification, in the context of chromosome dynamics, are listed in Table 7.1. These modifications are usually highly specific, and restricted mainly to the two exposed and flexible ends of each polypeptide histone chain. For example, acetylation, by reducing the net positive charge, reduces the strength of binding of those histones to the negatively-charged DNA. Indeed, histones carrying many acetyl groups have been known for a long time to be associated with chromatin that is being actively transcribed, or else easily available for transcription. Conversely, chromatin which is transcriptionally inactive or highly folded usually lacks acetyl groups. Similarly, phosphorylation of a particular serine residue (serine 10) in histone H3 is associated both with the condensation of metaphase chromosomes, and also with the activation of specific genes during interphase.

Histone methylation is generally associated with the recruitment of other proteins, which may then combine to form an inactive chromatin structure. Once formed, that inactive structure will remain stable to DNA replication and cell division; the underlying molecular basis for such remarkable stability will be described in Chapter 11. The most significant site of protein methylation seems to be located on lysine 9 of histone H3, where as tri-methyl-lysine it is recognized by another chromatin protein HP1, which has the ability to

repress transcription. In another case, the methylation of lysine 4, also on histone H3, is associated with the *activation* of genes.

Histone modifications such as acetylation are intimately associated with increased nucleosome accessibility. Nucleosomes can be moved along the DNA or 'shuffled' by large remodelling complexes, which were first characterized by Carl Wu. This process can be helped by multiple acetylation of histones; and in certain cases it may result in the complete displacement of the histone octamer from a promoter region, thereby allowing both transcription factors and RNA polymerase to bind to the DNA.

In the nucleus, one type of RNA polymerase, responsible for the synthesis of messenger-RNA, is also heavily involved in changing the properties of nucleosomes. The largest protein subunit of RNA polymerase possesses a long 'tail' which consists of 30 to 50 repeats of a seven-amino-acid sequence (tyrosine-serine-proline-threonine-serine-proline-serine, or YSPTSPS). When RNA polymerase binds to any gene promoter, but before it starts to make RNA, the long tail-repeat may attract an 'activator' complex of proteins which is required for the proper initiation of RNA chains: see Fig. 7.11(a). Then during the initiation process itself, that same polymerase tail will become phosphorylated on one of its three serines (number 5 of the repeat); and will bind to a different, large 'elongator' complex of proteins which assists in 'capping' of RNA chains at their 5′-ends, to protect them from cellular nucleases: see Fig. 7.11(b). During subsequent stages of transcription, the tail becomes phosphorylated on serine number 2 of the repeat (instead of number 5), as shown schematically in Fig. 7.11(c). Then it may bind also to other large enzymatic complexes which help to 'splice' the messenger-RNA (i.e. remove non-coding parts before it leaves the nucleus), or 'polyadenylate' the messenger-RNA (i.e. add a string of A bases at its 3′-end to protect against nucleases). One of the protein subunits of that elongator complex is actually an acetylase enzyme, which can modify the lysine side-chains of histones within active genes.

So the cell acts very cleverly to couple transcription of DNA into RNA by any polymerase enzyme, and its associated 'activator' or 'elongator' complexes, with acetylation of histone proteins so as to establish an active state.

Now what exactly might happen, when RNA polymerase approaches some nucleosome in the chromatin which is blocking its path? Often the DNA which lies between that advancing polymerase and the bound nucleosome will become supercoiled in a right-handed sense (see Chapter 6); which will facilitate the unwrapping or 'loosening' of a left-handed, two-turn supercoil of DNA within the nucleosome itself. One of the histone H2A-H2B dimers

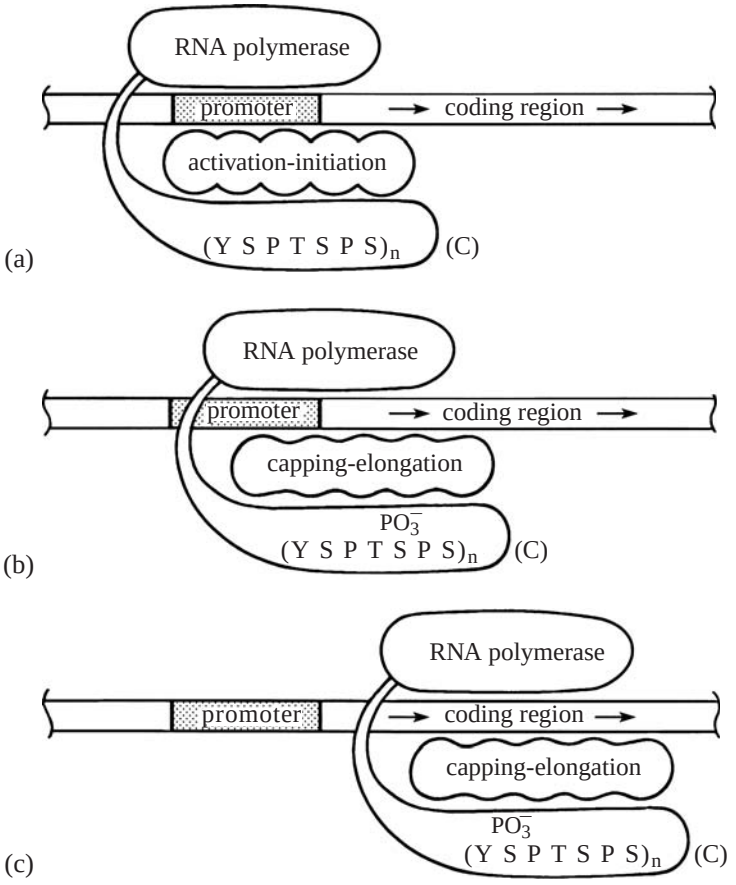

Figure 7.11 A highly schematic picture of RNA polymerase with its long 'tail' as it moves along the DNA. In (a), the tail has bound the 'activator' complex. Here, the seven-amino-acid repeating sequence is written in single-letter code: see the text. In (b), later on, one of the serines (S) in the tail repeat-region becomes phosphorylated; and a different, 'elongator' complex is bound instead. In (c), a different serine in the repeat has become phosphorylated; and now the complex bound to the tail includes a histone acetylase. Here, (C) indicates the carboxyl terminus of the protein.

may then be transiently released, as shown schematically in Fig. 7.12, thereby providing better access for the polymerase; although no one is sure how that might be accomplished. But another kind of RNA polymerase – polymerase III – appears to gain access to the DNA without displacement of histones from the octamer.

Over 20 years ago Brad Baer and Daniela Rhodes showed that some of the nucleosomes from actively growing cells lack one of those H2A-H2B dimers. Intriguingly, these particles were not only associated preferentially with RNA polymerase, but also were more accessible to nuclease cleavage than normal nucleosomes (i.e. they were unwrapped).

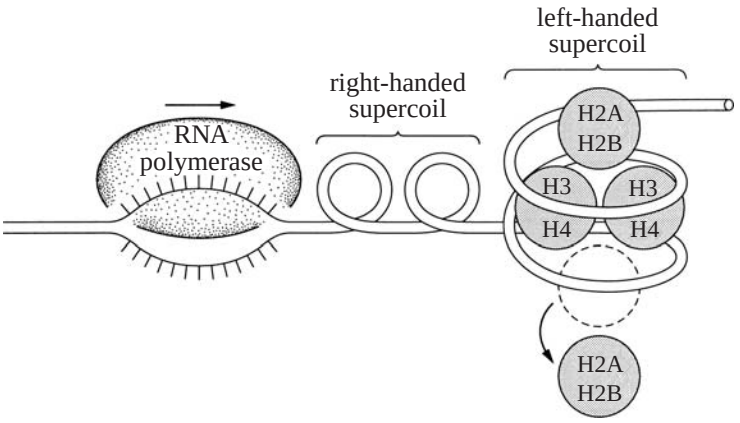

Figure 7.12 As the RNA polymerase advances (compare Fig. 4.2), right-handed supercoiling is imparted to the double-helical DNA ahead of it. This tends to loosen the left-handed supercoil of DNA around the histone octamer; and one of the H2A/H2B dimers may be expelled. Here, we have represented the histone octamer highly schematically as a bundle of four dimers.

Some nucleosome 'remodelling' enzymes seem to be attracted to regions of active chromatin; either by binding to acetylated histones, or else by binding to gene-activation proteins known as 'transcription factors' (see Chapter 8). Most 'remodelling' enzymes, like RNA polymerase, require an input of biochemical energy to move forward along any histone-bound DNA. Typically this comes from the conversion of adenosine tri-phosphate (ATP) to di-phosphate (ADP), by the replacement of a phosphate group with water and simultaneous release of energy (i.e. the three phosphates attached to an adenine base with a sugar ring are reduced to two; see Chapter 10). Moreover, most of those remodelling enzymes can also produce the transient release of H2A-H2B dimers, again just as for the action of RNA polymerase.

In summary, the acetylation of lysines in histones, plus the local supercoiling of DNA in a positive sense, as well as a transient release of H2A-H2B dimers, all combine to help RNA polymerase transcribe through any long array of nucleosomes without permanently displacing them.

Now we have covered many major aspects of current research into chromosome structure and gene activity. It is clear that the solution of some of the great problems in biology, such as the growth and development of higher organisms, will require a much deeper understanding of chromosomes and how they work.

Note

1. See Appendix 1.