Research on chromatin structure and function is expanding rapidly. Technical advances allow us to follow the events regulating gene expression in the eukaryotic nucleus in molecular detail. Within the chromosome, alterations in the organization and accessibility of key regulatory DNA sequences can be documented and interpreted. This book is intended to introduce scientists to this exciting field, in the expectation that many more contributions will be required before we understand completely how the nucleus of a eukaryotic cell functions.

The book has five sections. The first section is a brief overview of the issues discussed and an historical account of their development. The second section describes the structure of chromatin and chromosomes as far as it is known. Concepts concerning chromatin structure are already very well developed; indeed, many of the biophysical techniques and paradigms for studying protein–nucleic acid interactions were pioneered using the basic unit of chromatin, the nucleosome, as a model. In contrast, large-scale chromosomal architecture is much less well defined, as is the influence of modifications of structural proteins on chromatin and chromosome organization. How these changes may contribute to the various requirements for correct chromosomal function is a recurring theme.

A complete understanding of the eukaryotic nucleus requires not only that we know how to take it apart, but also that we can assemble it from the various component macromolecules. The third section describes the approaches, results and interpretations of experiments designed to accomplish this task. The biological constraints of
assembling a chromosome rapidly are discussed with reference to its final form and properties.

Form and function are intimately related. Once a complete understanding of a process is achieved, it is impossible to separate one from the other. The fourth section describes the multitude of approaches taken towards resolving how DNA can be folded into a chromosome and still remain accessible to the regulatory proteins, and allow processive enzymes to move along the length of the DNA molecules. It is in this field of research that much of the current progress on the interrelationship of chromatin structure and function is taking place. The final section offers a perspective on where prospects for future development might lie.

I would like to thank participants in the NIH chromatin group for sharing their ideas and results, especially Drs Trevor Archer, David Clarke and Sharon Roth. I am indebted to Drs Randall Morse, Geneviève Almouzni, Jeffrey Hayes and my Editor Dr Susan King for their comments on the text. Appreciation and thanks are given to Ms Thuy Vo and Mr William Mapes for preparing the manuscript and figures. Finally I thank my wife Elizabeth for her patience and support during the preparation of this book.

Alan Wolffe
The impact of chromatin structure on gene activity and many other nuclear events has become increasingly apparent over the past four years. Tremendous progress has been made concerning the structure and function of the nucleoprotein structures regulating transcription, replication and repair within the eukaryotic chromosome. Important recent advances include the determination of the internal organization of the nucleosome. The histones are found to have unexpected structural similarities to known transcription factors. Similar structures point to similar functions and this emphasizes the importance of considering both the architectural roles of histones and transcription factors in regulatory complexes. Genetic experiments have introduced a whole new significance both to the histones and to other proteins that control long-range chromosomal compaction and regulate differential gene activity. The current text has been extensively modified to incorporate such new discoveries into the framework of established knowledge. The principal aim remains to introduce interested scientists to chromatin.

I would like to thank my colleagues at NIH for sharing their ideas and results. I am indebted to Drs Dmitry Pruss, Horace Drew, Jeffrey Hayes, Stefan Dimitrov, Mary Dasso and Geneviève Almouzni for invaluable discussions. Drs Randall Morse and Jeffrey Hansen read the text for which I am particularly grateful. The interpretation of data
and any errors are my own. Appreciation and thanks are given to my Editor Dr Tessa Picknett and to Ms Thuy Vo for help with preparation of the text. Finally, I thank Elizabeth and Max for their patience and support.

Alan Wolffe
February 1995
Progress in chromatin research in the past three years has been remarkable. Pre-eminent in recent discoveries is the role of transcriptional coactivators and corepressors as histone modification enzymes. Scientists investigating transcriptional control and signal transduction are now faced with the need to consider chromatin structural modifications as a primary regulatory mechanism. Other advances concerning the nucleosome include the definition of unusual chromatin architecture on human disease genes, the expansion of the families of proteins that resemble chromatin components, and the solution of the crystal structure of the nucleosome core. The nucleus itself is also increasingly recognized as having structural and functional compartmentalization. This organization can contribute to epigenetic effects that have important roles in gene expression and development. The reversibility of such compartmentalization has been dramatically demonstrated through the successful mammalian cloning experiments. New sections and extensive rewriting have integrated these discoveries into the framework of established knowledge. The principal aim remains to introduce interested scientists to chromatin.

I would like to recognize the contributions of my colleagues at NIH and especially the Chromatin Interest Group in sharing their ideas and results. I am indebted to Drs Dmitry Pruss for many of the illustrations, and to Drs Mary Dasso, Jeffrey Hansen, Stefan Kass, Hitoshi Kurumizaka, Nicoletta Landsberger, Guofu Li, John Strouboulis, Alexander Strunnikov, Paul Wade and Jiemin Wong for
invaluable discussions. A special thanks to Ms Thuy Vo for the preparation of the text. I thank my wife Elizabeth for her patience and support, Max and Katherine for limiting their destruction of the manuscript.

I also thank Dr Tessa Picknett and Siân Davies at Academic Press, for their assistance in bringing this edition to print. Thanks also to Blackwell Science, the Company of Biologists, Elsevier Science and Oxford University Press for permission to reproduce previously published material.

Alan Wolffe
September 1997
CHAPTER ONE

Overview

1.1 INTRODUCTORY COMMENTS

Our knowledge of how the hereditary information within eukaryotic chromosomes is organized and used by a cell has increased enormously through the application of molecular biology and genetics. Technical advances now allow individual DNA sequences to be isolated and their association with proteins within the cell nucleus to be determined. Experimental progress has led the biologist to explore long-standing questions concerning how a particular cell acquires and maintains its individual identity. Developmental biologists have used new methodologies to investigate at a molecular level how an egg differentiates into different cell types. These questions have led scientists to the realization that growth, development and differentiation are directed by regulated changes in the form and composition of specific complexes of protein and DNA within the nucleus. Understanding how these complexes are assembled and function has become a central theme in modern biology.

Many of the techniques used to probe protein–DNA interactions were developed by researchers interested in the basic structural matrix of chromosomes – chromatin. This complex of DNA, histones and non-histone proteins has been exposed to a multitude of biochemical, biophysical, molecular biological and genetic manipulations. The structure of chromatin is by now well understood, but how it is folded and compacted into a chromosome is not. Knowledge of how
chromatin is constructed preceded the development of methods capable of exploring function. The purification and cloning of non-histone proteins required to perform the complex events involved in DNA transcription, replication, recombination and repair is the focus of a continuing and intense research effort. Investigators now make use of their experience with chromatin structure and assembly to examine the function of the structural proteins and enzymes required for the maintenance, expression and duplication of the genome in a true chromosomal environment.

The conclusion from this research effort is that the organization of DNA into chromatin and chromosomes is essential for regulated processes within the nucleus. Histones, nucleosomes and the chromatin structures they assemble function as integral components of the machinery determining transcriptional activity, cellular identity and fate. It might be anticipated that a comparable integration of structure and function will have occurred with the molecular machines controlling replication, recombination and repair.

1.2 DEVELOPMENT OF RESEARCH INTO CHROMATIN STRUCTURE AND FUNCTION

Towards the end of the nineteenth century numerous investigators formulated the theory that chromosomes determined inherited characteristics (see Voeller, 1968). These studies were almost entirely based on cytological observations with the light microscope. Although chromosomes are clearly only present in the nucleus, the influence of components of the cytoplasm on inherited characteristics was examined by forcing embryonic nuclei into regions of the cytoplasm in which they would not normally be found (Wilson, 1925). These experiments and others led Morgan (1934) to propose the theory that differentiation depended on variation in the activity of genes in different cell types. The genes were clearly in the chromosomes, but their biochemical composition remained completely unknown.

The last quarter of the nineteenth century also saw the recognition of RNA (first identified as yeast nucleic acid), DNA (thymus nucleic acid) and the discovery of histones. Albrecht Kossel isolated nuclei from the erythrocytes of geese and examined the basic proteins in his preparations, which he named the histones (reviewed by Kossel, 1928). The apparent biochemical simplicity of DNA and the obvious complexity of protein in chromosomes led investigators mistakenly to regard the latter component as the major constituent of the elusive
Overview

genes (Stedman and Stedman, 1947). Only the gradual acceptance of experiments on the capacity of DNA alone to change the genetic characteristics of the cell (Avery et al., 1944) led to the recognition of nucleic acid as the key structural component of a gene.

The elucidation of the double helical structure of DNA with its immediate implications for self-duplication, opened up the new approaches of molecular biology to clarifying the nature of genes (Watson and Crick, 1953). Although the double helix was now recognized as containing the requisite information to specify a genetic function, how this information was controlled was not understood. The apparent heterogeneity of the histones due to proteolysis and the various modifications of these proteins suggested that they might be important in regulating genes. Eventually methodological improvements for isolating and resolving the different histones demonstrated that they were highly conserved in eukaryotes and that only a few basic types existed (Fitzsimmons and Wolstenholme, 1976). This lack of variety implied that histones themselves were unlikely to be the determinants of gene specific transcription. However, a key role for histone modification remained central to prevailing ideas of transcriptional regulation (Allfrey et al., 1964).

A major breakthrough came in the 1970s when a combination of methodologies, including nuclease digestion, protein–protein cross-linking, electron microscopy and sedimentation analysis, determined that chromatin consisted of a repetitive fundamental nucleoprotein complex, which came to be called the nucleosome.

Structural studies on the nucleosome continue to the present time. Current and past research reveals the nucleosome to be a remarkably complex structure in which DNA is wrapped around the histones. The integrity of the nucleosome depends on highly specific histone–histone interactions, and the recognition by the histones of DNA structural features as the nucleosome is assembled. The core histones are present as an octamer, consisting of two molecules of H2A, H2B, H3 and H4. Histones H3 and H4 assemble a tetramer (H3, H4)2 that wraps DNA such that two dimers of H2A and H2B can stably associate. Once two turns of DNA are wrapped around the octamer, a fifth linker histone, such as histone H1, can be stably incorporated to complete the assembly process. Although all nucleosomes maintain these architectural features, there are many variations built upon this common theme.

Nucleosomal structures can contain different forms of particular core histones or linker histones. These histone variants are the products of distinct genes which may be differentially expressed during development (Newrock et al., 1977). The histones can also be post-
translationally modified to different extents. Early experiments associated different types of histone modification with particular nuclear functions such as transcription (Allfrey et al., 1964). Many early attempts were made to interrelate general differences in the transcriptional activity of genes to the solubility properties of chromatin dependent on histone modification or differences in histone content.

Recombinant DNA methodologies facilitated the isolation and cloning of defined DNA sequences, and DNA sequencing enabled the cis-acting elements potentially controlling gene expression to be defined (Brown, 1981). Hybridization analysis allowed the transcriptional activity of specific genes to be related to their accessibility to nucleases such as DNase I (Weintraub and Groudine, 1976). More detailed studies revealed that the regulatory DNA, such as promoter and enhancer sequences, was hypersensitive to DNase I cleavage (Wu et al., 1979). Chromatin was perceived as having a precise organization that was certainly modified by the transcription process. It was even possible to infer that structural features of chromatin might actually determine the potential for transcription to occur. Nevertheless, analysis of the nuclease sensitivity of chromatin was primarily descriptive. The molecules that actually directed the transcription of specific eukaryotic genes could not be determined through these approaches.

The enzymatic activities of the eukaryotic RNA polymerases had been characterized through the early 1970s. An initially disappointing conclusion from these studies was that these polymerases alone did not recognize the regulatory elements of eukaryotic genes with any specificity when the template was naked DNA. Roeder and colleagues (Parker and Roeder, 1977; Jaehning and Roeder, 1977) made the seminal discovery that RNA polymerases would accurately transcribe genes within chromatin, but not as naked DNA. The hunt was now on for the auxiliary proteins that would determine the specific initiation of transcription by RNA polymerase.

The early searches for these transcription factors were dependent on the development of in vivo and in vitro assays for transcription. Microinjection of purified or cloned genes into the nuclei of eukaryotic cells was an early assay system used to define the cis-acting sequences recognized by transcription factors (Brown and Gurdon, 1977). Subsequent assays relied on in vitro transcription extracts (Wu, 1978; Weil et al., 1979). These assays led to the purification and characterization of the first gene-specific eukaryotic transcription factor in 1980 (Engelke et al., 1980; Pelham and Brown, 1980).

Much of the research effort on transcriptional regulation during the 1980s focused on the further definition of cis-acting elements and
trans-acting factors involved in the initiation of the transcription process (Johnson and McKnight, 1989). The in vitro transcription or transfection assays used to examine the function of transcription factors did not require templates to be within their normal chromosomal environment for transcription to occur. In general these assays examined mechanisms that stimulated gene transcription, but did not examine the repression of transcription or the regulation of transcription in a physiological context.

Although far from the mainstream of research on transcription, the 1980s also saw the discovery of nucleosome positioning around eukaryotic genes (Simpson and Stafford, 1983). Application of genomic footprinting methodologies established that this phenomenon was a feature of several regulatory DNA sequences (Almer et al., 1986; Richard-Foy and Hager, 1987). Histones were increasingly perceived as having the potential for specific effects on the transcription process. Experiments that combined in vitro transcription systems with natural chromosomal templates revealed a specific role for histones in transcriptional regulation (Schlissel and Brown, 1984). All of this work relied upon the detailed analysis of particular promoters in individual laboratories. The overall relevance of chromatin structure to the eukaryotic transcription process was difficult to establish from these studies. Nevertheless, they provided the foundation for the interpretation of genetic experiments that did in fact determine the general significance for transcription of assembling DNA into nucleosomes.

In a series of insightful experiments Grunstein, Winston and colleagues (Han et al., 1987, 1988; Han and Grunstein, 1988; Clark-Adams et al., 1988) determined that changes in nucleosomal packaging had pleiotropic effects on gene activity. Subsequent work by these investigators and Mitch Smith established that very specific modifications in histone structure could either activate or repress specific genes (Megee et al., 1990; Durrin et al., 1991; Mann and Grunstein, 1992). This led directly to the resurgence of interest towards understanding gene activity in the natural chromosomal environment that has characterized much of the research effort in eukaryotic transcriptional regulation over the past few years.

The new-found interest in the role of chromatin in transcriptional regulation has been fuelled by progress in two specific areas. Structural studies led to the recognition that the histones were isomorphous with components of the transcriptional machinery (Arents and Moudrianakis, 1993; Clark et al., 1995; Ramakrishnan et al., 1993; Xie et al., 1996; Luger et al., 1997). These observations provided an architectural foundation for examining the specific roles
of histones and transcription factors in the assembly and function of regulatory nucleoprotein complexes. Specific modifications to nucleosomal architecture through histone acetylation, removal of histones H2A/H2B or H1 were shown to alleviate the repressive effects of chromatin assembly (Lee et al., 1993; Bouvet et al., 1994; Ura et al., 1995). In certain instances chromatin assembly was also shown to stimulate the transcription process (Schild et al., 1993). Thus the potential roles of nucleosomal proteins in gene control became more interesting (van Holde, 1993). Biochemical purification of histone acetyltransferases and deacetylases (Brownell et al., 1996; Taunton et al., 1996) provided an even closer link between chromatin and the transcriptional machinery. Histone acetyltransferases were discovered to be components of large macromolecular complexes known as coactivators, which are targeted to specific promoters by transcriptional activators. Therefore a direct link was established between histone acetylation and transcriptional activation. Histone deacetylases were found within corepressor complexes that turn genes off. Once again, histone chemistry became an important variable to consider in transcriptional control.

It is now recognized that to understand transcriptional control or any other regulated event in the nucleus it is necessary to define the chromatin structure within which DNA is utilized. Aside from the characterization of specific architecture, we must also determine how structure might change. Chromatin is not static, but dynamic. Targeted histone modifications within regulatory nucleoprotein complexes have emerged as a means of modulating the stability of repressive chromatin structures and the transcription process itself. The observations made using simple model systems are having an impact on our understanding of both development and disease. It is now probable that our increasing knowledge of both chromatin and chromosome structure and function in the nucleus will provide many avenues for future advances in biotechnological and medical fields.
Chromatin Structure

Chromosomes represent the largest and most visible physical structures involved in the transfer of genetic information. Surprisingly, our understanding of chromosome organization is most complete for the smallest and most fundamental structural units. These units are the nucleosomes which contain both DNA and histones. Long folded arrays of nucleosomes comprise the vast majority of chromatin. In this section I discuss the structural features of DNA and histones, how they assemble into nucleosomes and how nucleosomes fold into chromatin fibres. Finally, I describe what we know about the organization of the chromatin fibre into a chromosome and how this can be modified in various ways.

2.1 DNA AND HISTONES

The most striking property of a chromosome is the length of each molecule of DNA incorporated and folded into it. The human genome of $3 \times 10^9$ bp would extend over a metre if unravelled; however, this is compacted into a nucleus of only $10^{-5}$ m in diameter. It is an astonishing feat of engineering to organize the long linear DNA molecule within ordered structures that can reversibly fold and unfold within the chromosome. Not surprisingly, many aspects of chromosome structure reflect the impediments and constraints imposed by having to bend and distort DNA.
2.1.1 DNA structure

DNA has an elegant and simple structure around which the chromosome is assembled. The DNA molecule exists as a long unbranched double helix consisting of two antiparallel polynucleotide chains. DNA always contains an equivalent amount of the deoxyribonucleotide containing the base adenine (A) to that with the base thymine (T), and likewise of the deoxyribonucleotide containing the base guanine (G) to that with the base cytosine (C) (Fig. 2.1). Each base is linked to the pentose sugar ring (2-deoxyribose) and a phosphate group. The 5' position of one pentose ring is connected to the 3' position of the next pentose ring via the phosphate group (a 5'-3' linkage) to create the polynucleotide chain (Fig. 2.2). The two antiparallel polynucleotide chains are attached to each other by hydrogen bonding between the bases. G is always base paired to C, and A is always base paired to T. In addition to the stability imparted by hydrogen bonding, hydrophobic base stacking interactions occur along the middle of the double helix (Fig. 2.3) (see Calladine and Drew, 1997 or Sinden, 1994 for details).

Physical studies using X-ray diffraction indicate that under conditions of physiological ionic strength, DNA is a regular helix, making a complete turn every 3.4 nm with a diameter of 2 nm. This particular DNA structure is known as B-DNA and has approximately 10.5 bp/turn of the helix. This means that every base pair is rotated

Figure 2.1. The four bases found in DNA.
Figure 2.2. A nucleotide and a polynucleotide chain.

approximately 34° around the axis of the helix relative to the next base pair. This results in a twisting of the two polynucleotide strands around each other. A double helix is formed that has a minor groove (approximately 1.2 nm across) and a major groove (approximately 2.2 nm across). The geometry of the major and minor grooves of DNA will be seen later to be crucial in determining the interaction of proteins with the DNA backbone. The double helix is right handed (Fig. 2.4).

Beyond this basic description, DNA structure is exceedingly plastic. Crystallization of various oligonucleotides indicates that a variety of DNA sequences will yield recognizable B-form DNA structures (Privé et al., 1991; Yanagi et al., 1991). More severe alterations in the conditions under which DNA is examined do, however, generate
distinct conformations. Dehydrating the fibre will cause the double helix to take up a structure known as A-DNA (11 bp/turn); or placing DNA with a defined sequence of alternating G and C bases in solutions of high ionic strength will lead to the formation of a left-handed helix known as Z-DNA (12 bp/turn). The existence of either of these extreme structures in the eukaryotic nucleus under normal physiological conditions is controversial. However, their formation indicates the gross morphological changes that DNA can be forced to undergo (Drew et al., 1988; Calladine and Drew, 1997).

How do we know what structure populations of DNA molecules have in solution? Two experimental methodologies have been commonly used. The first employs DNA cleavage reagents and a flat crystal surface (Rhodes and Klug, 1980). When DNA is absorbed from solution on to a flat calcium phosphate surface and cut with DNase I, the enzyme cuts DNA most readily where it is exposed away from the surface. The average spacing between the sites of cleavage gives the approximate number of base pairs per turn of DNA (Fig. 2.5). This is
determined by the electrophoresis of denatured molecules through a polyacrylamide gel. A better reagent for this purpose is the hydroxyl radical. Hydroxyl radicals are generated by the Fenton reaction in which an Fe(II) EDTA complex reduces hydrogen peroxide to a hydroxide anion and a hydroxyl radical.

\[
\text{[Fe(EDTA)]}^{2-} + \text{H}_2\text{O}_2 \rightarrow \text{[Fe(EDTA)]}^{1-} + \text{OH}^- + \cdot\text{OH}
\]

The radical is about the size of a water molecule and has little sequence specificity in cleaving DNA. This it does by breaking the pentose sugar rings of individual deoxyribonucleotides. In contrast, DNase I is a large enzyme which has considerable sequence preferences. In both instances, the number of base pairs per turn of a large population of different DNA sequences bound to a crystal surface is found to be 10.5 (Tullius and Dombroski, 1985). This result is consistent with DNA having a B-form configuration as determined by X-ray studies.

The second method to examine DNA structure in solution reaches the similar conclusion that DNA has a B-form conformation at physiological ionic strength; however, a completely different strategy is used. It is generally found that a population of closed circular DNA

---

**Figure 2.4.** The dimensions of DNA. Base pairs are shown as horizontal lines for one turn of the double helix.
Figure 2.5. Determining the helical periodicity of DNA in 'solution' through binding to a flat crystal surface and cleavage with an enzyme or a chemical reagent. In theory the most exposed region of the double helix will be cut preferentially, experimentally this is reflected in a larger population of DNA fragments cut at this site after resolution on a polyacrylamide gel (darker bands). The distance between darker bands in base pairs is the helical periodicity (number of base pairs per turn) of DNA.

molecules, identical in length and sequence, contains different numbers of superhelical turns. Superhelical turns can be simply defined by the following description: a single superhelical turn is introduced into a closed circular DNA molecule if the molecule is broken, one end of the molecule is then fixed, the other is rotated once and the two ends then rejoined. Supercoils can be positive or negative depending on which way the free DNA end is rotated. Closed circular molecules of the same length and sequence with different numbers of superhelical turns are known as topoisomers. Each population of small closed circular DNA molecules that differ in length by a few base pairs will exist as a distribution of topoisomers. These can be resolved by electrophoresis through an agarose gel matrix. A molecule which has a length corresponding to an integral number of helical turns will exist predominantly as a single topoisomer whereas a
molecule which deviates from this by half a helical turn will be equally likely to exist with the superhelical turn in a positive or negative sense. The number of DNA molecules with a particular mobility in the agarose gel will be reduced by half since the molecules exist as an equal mixture of topoisomers. Examining the relationship between DNA length and the distribution of topoisomers allows the number of base pairs per turn of DNA to be calculated. The result of 10.5 bp/turn is close to that derived from crystal binding studies (Horowitz and Wang, 1984). Finally, theoretical calculations of the most stable configuration of DNA, which actually preceded much of the experimental work, suggested a value of 10.6 bp/turn (Levitt, 1978). The range of values around 10.5 bp/turn, obtained both experimentally and theoretically, provides a sound basis for considering alterations in this structure based on DNA sequence content and histone–DNA interaction.

Aside from the dramatic changes in DNA structure seen on formation of A- or Z-DNA, local variations in DNA sequence can significantly influence DNA conformation and properties of the helix. Our most extensive knowledge of the local changes in B-form DNA structure due to sequence content comes from studying AT-rich DNAs. For example, oligo(dA).oligo(dT) tracts are found experimentally, using both spectroscopic techniques and DNA cleavage reagents such as the hydroxyl radical, to be straight and rigid with a constant narrow minor groove width (Nelson et al., 1987; Hayes et al., 1991a). This is believed to be a consequence of maximizing the hydrophobic base stacking interactions between adjacent A.T base pairs in the DNA helix (Fig. 2.3). This stabilization process requires the bases to be more twisted relative to each other than would normally be found in typical B-form DNA. Chains of these base pairs have the correct geometry to allow at least two water molecules per base pair to become highly ordered along the DNA backbone. This creates a 'spine of hydration' which contributes to the rigidity of oligo(dA).oligo(dT) tracts (Berman, 1991). Changes in sequence that affect these structural features lead to widening of the minor groove; for example, a G.C base pair will disrupt the straight path and rigidity of an oligo(dA).oligo(dT) tract. In contrast to oligo(dA).oligo(dT), oligo [d(AT)] tracts are conformationally flexible. This flexibility is a consequence of not being able to achieve efficient hydrophobic base stacking interactions between consecutive T.A and A.T base pairs without severely distorting the DNA helix (Travers and Klug, 1987; Travers, 1989). Finally, short oligo(dA).oligo(dT) tracts (4–6 bp in length) that are phased with a periodicity similar to that of the DNA helix itself will cause the molecule to be curved. This is due to a
narrowing of the minor groove every turn of DNA caused by the phased oligo(dA).oligo(dT) tract (Koo et al., 1986). Periodicities that are greater or smaller than 10–11 bp will cause the normally straight DNA to take on a ‘corkscrew-like’ path. In spite of this wide variation in ‘B-form’ DNA structure, all of these DNA sequences can be assembled into chromatin (Section 2.2.5).

DNA structure is thought to have an important role in certain human genetic diseases characterized by the presence of repeats of particular trinucleotide sequences. These trinucleotide repeats are found in the gene whose aberrant expression leads to the disease phenotype (Bates and Lehrach, 1994; Sutherland and Richards, 1995). The segments of DNA containing trinucleotide repeats are unstable with the potential to expand from generation to generation. Two trinucleotide repeat sequences are of particular interest: 

(CTG)\textsubscript{n} is associated with many diseases including Huntington’s disease, myotonic dystrophy, spinocerebellar ataxia type 1, and hereditary dentatorubral-pallidoluysian atrophy; (CCG)\textsubscript{n} is associated with fragile X mental retardation. Normal individuals have relatively few copies of these repeat sequences whereas diseased individuals have many copies (> 50). The number of repeats influences both the expansion process and the disease. How this influence is exerted has been the focus of a great deal of attention (see also Section 2.2.5).

Of all the many potential trinucleotides present in the genome only reiterated CTG and CCG sequences show the special properties of instability and tendency towards expansion (Han et al., 1994). These sequences have the capacity to form stable hairpin structures when they reach a certain threshold length of 40 to 50 repeats (Gacy et al., 1995). It has been suggested that the ability to form stable hairpins might explain both the dependence on particular trinucleotides and the length of sequence for repeat expansion. The favoured model for expansion predicts that DNA polymerase might ‘slip’ on reiterated sequences during replication leading to small increases in trinucleotide repeat copy number. Once the copy number becomes large enough, the single-stranded DNA at the replication fork might form a stable hairpin looping out intervening DNA and leading a small ‘slip’ to generate a large expansion of the trinucleotide repeat sequence (Gacy et al., 1995). Under these special circumstances the capacity to form unusual hairpin DNA structures might contribute to the generation of a disease phenotype.

The (CTG)\textsubscript{n} and (CCG)\textsubscript{n} sequences appear to have no reason to adopt unusual structural features when present as duplex DNA, however there is evidence that these sequences might differ from