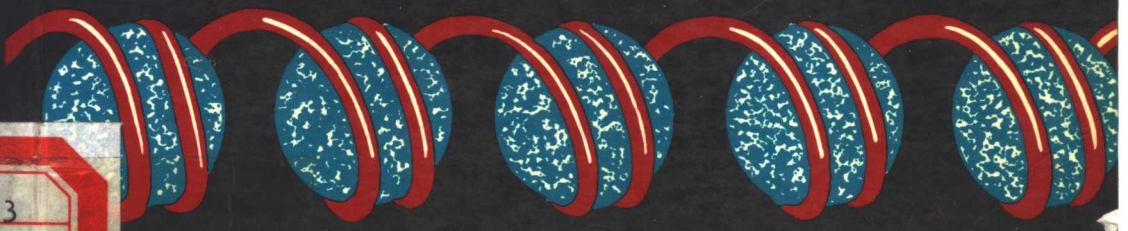
CHROMATIN

STRUCTURE & FUNCTION

Alan Wolffe



Chromatin

Structure and Function

A. WOLFFE

National Institutes of Health,
Bethesda, Maryland, USA

ACADEMIC PRESS
Harcourt Brace Jovanovich, Publishers
London San Diego New York
Boston Sydney Tokyo Toronto

This book is printed on acid-free paper

ACADEMIC PRESS LIMITED
24–28 Oval Road
London NW1 7DX

United States Edition published by
ACADEMIC PRESS INC.
San Diego, CA 92101

**A catalogue record for this book is available
from the British Library**

ISBN 0–12–761910–0
0–12–761911–9 (pbk.)

Typeset by Fakenham Photosetting Limited,
Fakenham, Norfolk
Printed in Great Britain by the University Press,
Cambridge

Preface

Research on chromatin structure and function is rapidly expanding. 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 chapters. Chapter 1 gives a brief overview of the issues discussed and an historical account of their development. In Chapter 2 the structure of chromatin and chromosomes is described 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 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. Chapter 3 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. Chapter 4 describes the multitude of approaches taken towards resolving how DNA can be folded into a chromosome and yet still remain accessible to the regulatory proteins, and allow processive enzymes to move along the length of the DNA molecule. 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 chapter 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 Clark 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

Contents

CHAPTER ONE

Introduction

Our knowledge of how the hereditary information within eukaryotic chromosomes is organized and used by a cell has increased enormously through the application of recombinant DNA methodologies. 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 the techniques of molecular biology and genetics to investigate how an egg differentiates into different cell types. These questions have led scientists to the realization that growth, development and differentiation proceed through 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 have the tools to try to integrate our experience with chromatin structure and assembly with the structural proteins and enzymes required for the maintenance, expression, and duplication of the chromosome.

1.1 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 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 primary regulators of gene activity.

During the 1970s the organization of the fundamental complex of DNA and histones, which came to be called the nucleosome, was largely solved. 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). Elucidation of the organization of regulatory DNA was immediately followed during the 1980s by the vigorous search for the non-histone proteins – the *trans*-acting factors that might interact and function at these regulatory elements. Many of these proteins have now been characterized (Johnson and McKnight, 1989). This endeavor has led to the current focus of molecular biology and genetics: the attempt to understand how *trans*-acting factors and the enzyme complexes involved in DNA replication, transcription, recombination and repair function *in vivo*. The continual improvement in experimental techniques has led to the realization that important regulatory elements in DNA are organized into specific structures including both the histone proteins and *trans*-acting factors. Much current research concerns this interrelationship, the structure of these complexes and their function in the utilization of the hereditary information within the chromosomes.

CHAPTER TWO

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. The long folded arrays of nucleosomes along the axis of a chromosome comprise the vast majority of chromatin. In this section we will discuss the structural features of DNA and histones, how they assemble into nucleosomes and how nucleosomes fold into chromatin fibers. Finally I will describe what we know about the organization of the chromatin fiber 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 3×10^9 bp would extend over a meter 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

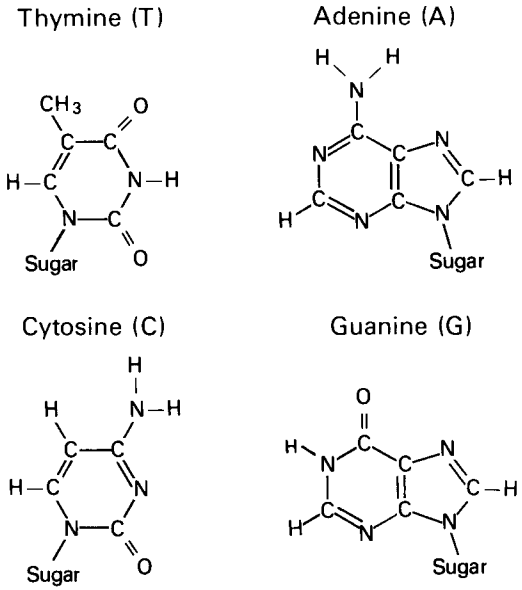


Figure 2.1. The four bases found in DNA.

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 Lewin, 1990 or Alberts *et al.*, 1990 for details).

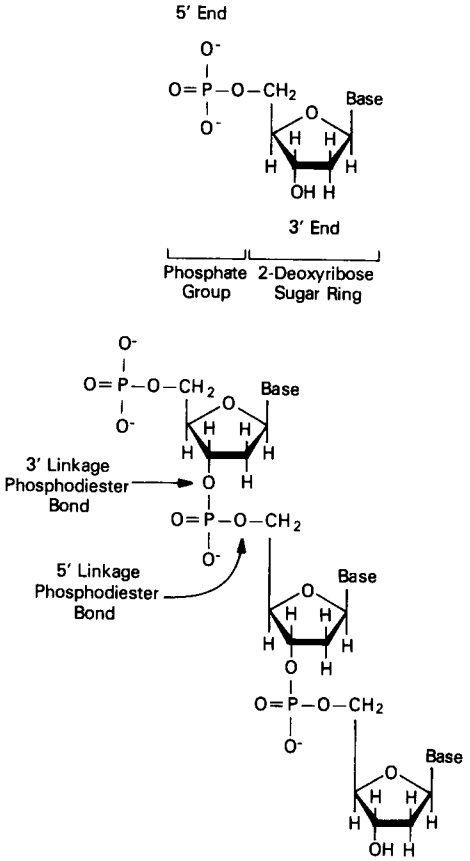


Figure 2.2. A nucleotide and a polynucleotide chain.

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 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.

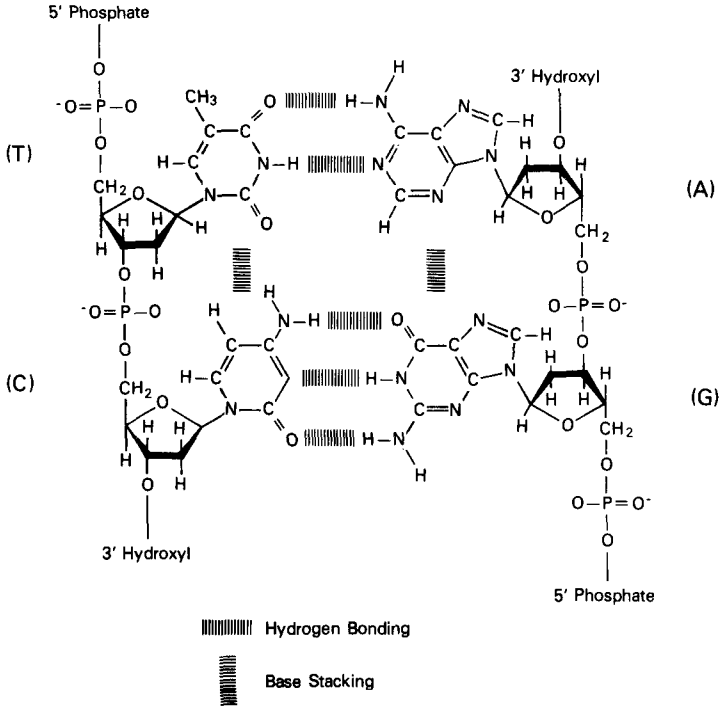


Figure 2.3. The interactions stabilizing the two antiparallel polynucleotide chains in DNA.

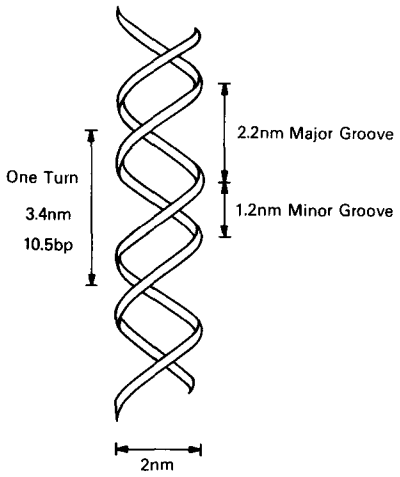


Figure 2.4. The dimensions of DNA.

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 fiber 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).

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 onto a flat calcium phosphate surface and cut with DNaseI, 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. 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, DNaseI 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 molecules, identical in length and sequence contains different numbers of superhelical turns. The formation of superhelical turns can be simply described as follows: 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

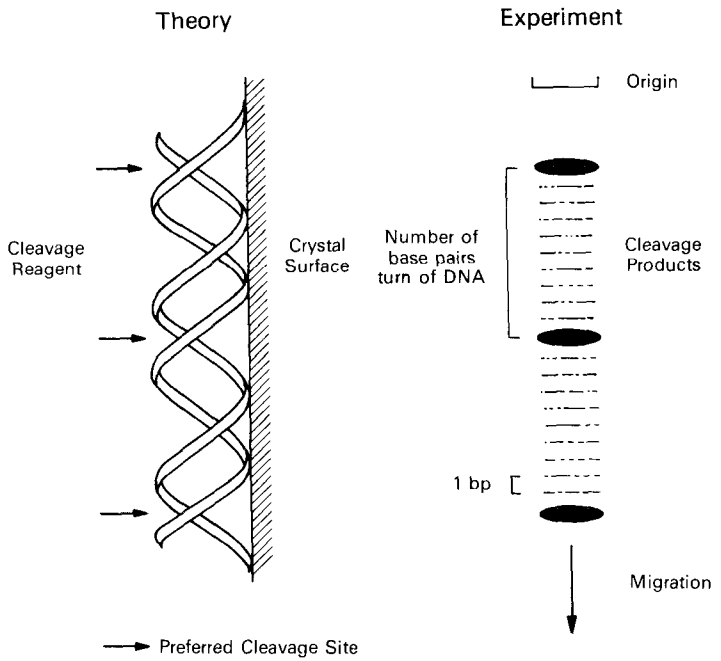


Figure 2.5. Determining the helical periodicity of DNA in 'solution' through binding to a flat crystal surface and cleavage with an enzyme or a chemical reagent.
In theory the most exposed region of the double helix will be cut preferentially, experimentally this is reflected in a larger population of 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.

the two ends 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