Advances in
PROTEIN CHEMISTRY

VOLUME 34

ADVANCES IN PROTEIN CHEMISTRY

EDITED BY

C. B. ANFINSEN

National Institute of Arthritis, Metabolism and Digestive Diseases Bethesda, Maryland

JOHN T. EDSALL

Biological Laboratories Harvard University Cambridge, Massachusetts

FREDERIC M. RICHARDS

Department of Molecular Biophysics and Biochemistry Yale University New Haven, Connecticut

VOLUME 34

1981

ACADEMIC PRESS

A Subsidiary of Harcourt Brace Jovanovich, Publishers

New York London Toronto Sydney San Francisco

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
24/28 Oval Road, London NW1 7DX

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 44–8853

ISBN 0–12–034234–0

PRINTED IN THE UNITED STATES OF AMERICA

81 82 83 84 9 8 7 6 5 4 3 2 1

ADVANCES IN PROTEIN CHEMISTRY

Volume 34

CONTRIBUTORS TO THIS VOLUME

Eddie Morild

Jane S. Richardson

Ruth Sperling

Ellen J. Wachtel

Donald B. Wetlaufer

CONTRIBUTORS TO VOLUME 34

Numbers in parentheses indicate the pages on which the authors' contributions begin.

EDDIE MORILD,* *Norwegian Underwater Institute, Gravdalsveien 255, N-5034 Ytre Laksevåg/Bergen, Norway* (93)

JANE S. RICHARDSON,† *Department of Anatomy, Duke University, Durham, North Carolina 27710* (167)

RUTH SPERLING,‡ *Department of Chemical Physics, Weizmann Institute of Science, Rehovot, Israel* (1)

ELLEN J. WACHTEL, *Department of Chemical Physics, Weizmann Institute of Science, Rehovot, Israel* (1)

DONALD B. WETLAUFER, *Department of Chemistry, University of Delaware, Newark, Delaware 19711* (61)

* Also at the Department of Chemistry, University of Bergen, N-5014 Bergen, Norway.

† Mailing address: 213 Medical Sciences IA, Duke University, Durham, North Carolina 27710.

‡ Present address: Department of Structural Biology, Stanford University School of Medicine, Stanford, California 94305.

ERRATUM

Advances in Protein Chemistry

Volume 33

Page 252 The text *should read*

of the peptide in the cerebellum. Apparently, the largely motor functions of the cerebellum do not use peptides as neurotransmitters. Cerebellar functions are basic ones which may have evolved before the neuropeptides made their appearance. The cerebral cortex also usually contains relatively low concentrations of neuropeptides, although its large size, in higher animals, makes this brain region a relatively rich source when the total

CONTENTS

The Histones

Folding of Protein Fragments

The Theory of Pressure Effects on Enzymes

The Anatomy and Taxonomy of Protein Structure

JANE S. RICHARDSON

THE HISTONES

By RUTH SPERLING[1] and ELLEN J. WACHTEL

Department of Chemical Physics, Weizmann Institute of Science, Rehovot, Israel

I. Introduction

A. Aim of the Article and Summary

The aim of this article is to present evidence and to develop a hypothesis concerning the dynamic structural role of histones in chromatin.

In higher organisms, unlike bacteria, the genetic information is stored in the nucleus. Duplex DNA is found complexed with an

[1] Present address: Department of Structural Biology, Stanford University School of Medicine, Stanford, California.

ADVANCES IN PROTEIN CHEMISTRY, Vol. 34

ISBN 0-12-034234-0

equal weight of basic proteins, the histones, and with a smaller amount of nonhistone proteins and RNA. This complex forms a highly compact, organized structure—the chromosome. During interphase, chromosomes expand, revealing the constituent structure called chromatin.

The current view of the histones is that they are primarily structural proteins. Independent of species and tissue, they interact in the same repetitive fashion with the chromosomal DNA to form nucleosomes. Their lack of diversity and non-sequence-specific binding to DNA argue against their playing any explicit role in gene regulation. On the other hand, histone microheterogeneity, arising from post-synthetic modifications and intraspecies sequence variations, still leaves open the possibility that histones are involved, in perhaps a subtle way, in the dynamics of chromosome structure.

The purpose of this article is to correlate the rather unique structural aspects of the five histone molecules—the differences among them as well as their similarities—with their biological function. Such an analysis is best approached, we believe, via a study of the many protein–protein (Section II) and protein–DNA (Section III) interactions in which the histones participate. Emphasis will be placed on the four core histones H2A, H2B, H3, and H4; H1 will be discussed briefly, mainly in relation to its interaction with DNA. In no sense is the bibliography meant to be exhaustive.

During the course of the article, we will present evidence that the histones constitute part of a self-assembly system (Section IV). In the absence of DNA they retain the information to interact with each other to form a hierarchy of structures with dimensions, periodicities, and intermolecular contacts, compatible with what is known about the protein core of chromatin. This strongly suggests that histone–histone interactions have a fundamental role in chromatin structure.

Evidence will be presented in Sections II and IV showing that, under conditions of high salt concentration, histone complexes related to the histone octamer, which is the histone core of the nucleosome, are obtained. Furthermore, higher oligomeric structures are obtained from complexes of H2A·H2B, H3·H4, and all four histones (acid- and salt-extracted). In addition, high ionic strength, while promoting the correct folding and complexing of the core histones, also promotes well-defined reversible histone helical fibers which have axial periodicities similar to those of chromatin. These fibers are formed by assembly of the histone tetrameric and octameric units. The arrangement of the histones in the histone fibers, on the one hand, is related to histone–histone interactions within the nucleosome, and, on the

other hand, it also suggests possible histone–histone interactions between nucleosomes. The intranucleosomal interactions of the histones are essential for the formation of the nucleosome. This refers not only to the histone arrangement within the protein core but also to the determination of the periodicity of the DNA fold around it. The internucleosomal interactions may bear relevance to chromatin function.

The similarity of the various histone fibers is probably correlated with the similarity in the distribution of the amino acids in the sequences of the four core histones and reflects their function as the skeleton or backbone of chromatin. However, from the presence of a specific pattern of interactions of the core histones and the existence of histone variants and histone postsynthetic modifications, one can anticipate modulations in the basic general pattern of histone structure. In Section V, a possible mechanism for histone microheterogeneity influencing chromatin structure is suggested. Analogous to other assembly systems, small subunit modifications may be amplified to produce major changes in the assembled superstructure.

For the benefit of those unfamiliar with the field of chromatin research this section continues with background material, characterizing components of the chromatin system which will be pertinent to our discussion.

B. Nucleosomes

It has been shown by biochemical and physical methods that chromatin is built of a repeating unit, the nucleosome (Kornberg, 1974; Olins and Olins, 1974; Woodcock, 1973; Sahasrabuddhe and Van Holde, 1974; Hewish and Burgoyne, 1973; Noll, 1974a; Oudet *et al.*, 1975; Baldwin *et al.*, 1975). This topic has been extensively reviewed (Kornberg, 1977; Felsenfeld, 1978; Chambon, 1978), and therefore only a brief summary will be presented here.

The nucleosome is composed of ~200 base pairs of DNA and an octamer of the histones H2A, H2B, H3, and H4 as well as histone H1 (Kornberg, 1974, 1977). Nucleosomes can be obtained by mild digestion of chromatin with micrococcal nuclease (Noll, 1974a; Axel, 1975), followed by fractionation on a sucrose gradient. Further digestion of the nucleosomes results in the formation of nucleosome core particles composed of 145 base pairs of DNA and an octamer of the histones H2A, H2B, H3, and H4 (Rill and Van Holde, 1973; Sollner-Webb and Felsenfeld, 1975; Axel, 1975; Bakayev *et al.*, 1975; Whitlock and Simpson, 1976; Noll and Kornberg, 1977). The DNA piece thus excised is called "linker" DNA which serves as a link

between nucleosomes. It was suggested that histone H1 is bound to the DNA linker (Whitlock and Simpson, 1976; Shaw *et al.*, 1976; Varshavsky *et al.*, 1976; Noll and Kornberg, 1977).

The length of the linker DNA has been shown to vary, resulting in nucleosomes with DNA lengths varying from 154 base pairs in *Aspergillus* to 241 base pairs in sea urchin sperm (for references, see Kornberg, 1977; Chambon, 1978). There are variations among different animal species and also among different cell types in the same organism. Only the linker DNA is variable, while the core DNA length is constant. Changes in H1 have been related to linker variability (Noll, 1976; Morris, 1976a,b).

It has been suggested by Kornberg (1974) that each nucleosome contains two copies of each of the four core histones H2A, H2B, H3, and H4. Immunological studies have shown that histone H2B is present in every nucleosome of rat liver and calf thymus chromatin (Bustin *et al.*, 1976). It was further shown that nucleosomes from HeLa cells contain all four histones (Bustin *et al.*, 1977). This, together with the demonstration of the existence of the histone octamer by cross-linking of chromatin (Thomas and Kornberg, 1975a) and measurements of histone-to-histone (Albright *et al.*, 1979) and histone-to-DNA ratios (Olins *et al.*, 1976; Joffe *et al.*, 1977), supports the model proposed by Kornberg (1974). X-Ray and neutron scattering studies have shown that chromatin has an organized structure with dominant periodicities of 110, 55, 37, and 27 Å (Luzzati and Nicolaieff, 1959; Wilkins *et al.*, 1959; Pardon *et al.*, 1967, 1975; Pardon and Wilkins, 1972; Baldwin *et al.*, 1975). The neutron experiments furthermore determined that the bulk of the histone proteins occupies the central region of the complex to a radius of approximately 32 Å, forming an inner core, and that the DNA is supercoiled on the outside (Baldwin *et al.*, 1975). The packing of the DNA with the histones to form nucleosomes results in an approximate sevenfold increase in the linear packing density of the DNA (Oudet *et al.*, 1975).

Recently, a low-resolution model of the chromatin "core" particle has been derived from a combination of single-crystal X-ray diffraction and electron microscopic data (Finch *et al.*, 1977). The particle is described as a flat cylinder 110 Å in diameter and 57 Å in height. A similar shape and similar dimensions were found to be consistent with the low-angle neutron scattering from core particles in solution (Pardon *et al.*, 1977; Suau *et al.*, 1977). Some conclusions may be drawn concerning the conformation of the DNA. Presumably, the strong 28 Å periodicity apparent in the crystal data (Finch *et al.*, 1977) corresponds to the pitch of the DNA superhelix wound about the histone core. X-Ray and spectroscopic data suggest that the DNA super-

coiled around the histone core remains predominantly in the B-conformation (Bram, 1971; Goodwin and Brahms, 1978). The radius of the superhelix must be about 45 Å which, for 145 base pairs of B-form DNA, corresponds to somewhat less than two turns. (1.75 turns have been suggested by Finch *et al.*, 1977). Very little can be deduced from the crystal data concerning the organization of histones in the crystals. However, there is a large body of indirect evidence concerning histone structure and interactions which will be discussed in later sections.

Concerning the path of the DNA around the histone core, a few models have been suggested. Theoretical calculations have shown that the DNA can bend smoothly around the histone core (Levitt, 1978; Sussman and Trifonov, 1978). However, a small change in the number of base pairs per turn relative to that in solution was proposed (Levitt, 1978). A different model was suggested in which the DNA kinks at every tenth or twentieth base pair (Crick and Klug, 1975; Sobell *et al.*, 1976).

Detailed digestion studies of nucleosomes and core particles have shown that a ladder of DNA fragments is obtained, indicating a regular periodic structure of the core particles. The "10-base-pair" repeat obtained by DNase I (Noll, 1974b) digestion of nucleosomes as well as by other nucleases (Sollner-Webb *et al.*, 1978) was recently determined more accurately to be a 10.4-base-pair repeat (Prunell *et al.*, 1979). This may reflect either the periodicity of the DNA double helix in chromatin or the DNA kinking. Experiments with 5′-end labeling of the DNA with ^{32}P followed by DNase I digestion have suggested a more complex structure, in which not all the sites have the same susceptibility for DNase I digestion. Sites at positions 30, 60, and 80 show significantly decreased susceptibility (Simpson and Whitlock, 1976; Lutter, 1978; Noll, 1978). It has been suggested (Finch *et al.*, 1977) that the less susceptible sites for DNase I are the binding sites of the histones which inhibit the digestion.

C. Higher Order Structure of Chromatin

The nucleosome is the building block of chromatin. Chromatin is a dynamic structure and it goes through several levels of compaction, reaching the very compact structure of mitotic chromosomes.

Although the higher order structure of chromatin is still not well understood, two models have been suggested and supported by experimental evidence (for references, see Felsenfeld, 1978; Chambon, 1978). One is the solenoid model of Finch and Klug (1976), obtained by supercoiling the chromatin thread, and the other model is the superbead model (Renz *et al.*, 1977). Each may represent a different

state of the same material. Evidence in the literature has suggested that histone H1 is involved in the higher order structure of chromatin (for references, see Felsenfeld, 1978; Chambon, 1978), and some models have been proposed to account for that possibility (Finch and Klug, 1976; Worcel, 1978).

D. The Histones

Since their discovery in 1884 by Kossel, the histones have been the subject of comprehensive study. This work has been summarized in several books and reviews (Phillips, 1971; DeLange and Smith, 1971, 1972; Hnilica, 1972; Elgin and Weintraub, 1975; Li, 1977; Isenberg, 1978, 1979; Von Holt *et al.*, 1979). Therefore, no attempt will be made here to give a comprehensive treatment of the work already reviewed. However, a brief discussion of some special properties of the histones will be presented before we proceed to the treatment of those structural aspects of the histones which are the focus of this article.

Concerning their primary structure, the histones can be divided into three groups: (1) very lysine-rich histone H1, (2) lysine-rich his-

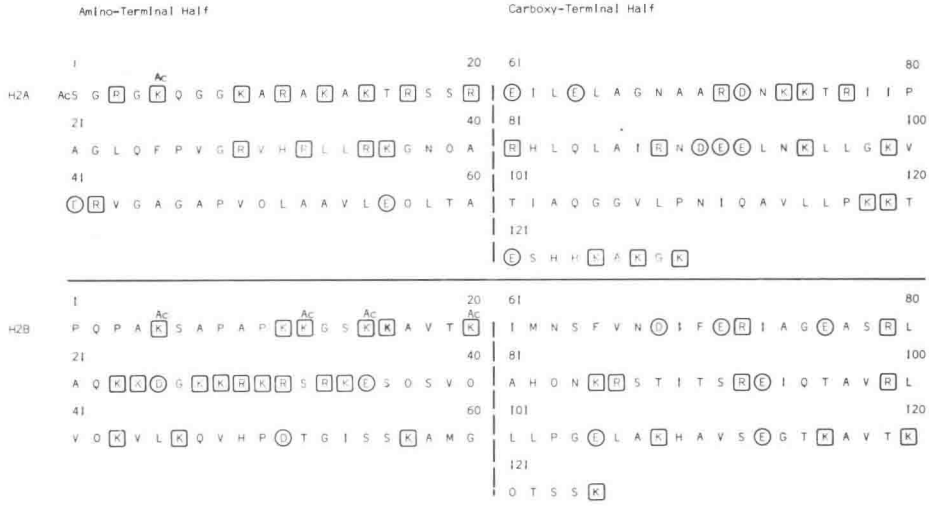

FIG. 1. The amino acid sequences of calf histones H2A (Yeoman *et al.*, 1972; Sautiere *et al.*, 1974) and H2B (Iwai *et al.*, 1972). A one-letter code is used: A, alanine; R, arginine; N, asparagine; D, aspartic acid; C, cysteine; E, glutamic acid; Q, glutamine; G, glycine; H, histidine; I, isoleucine; L, leucine; K, lysine; M, methionine; F, phenylalanine; P, proline; S, serine; T, threonine; O, tyrosine; V, valine. Basic residues are indictated by squares and acidic residues are indicated by circles. Ac denotes acetylation site.

tones H2A and H2B, and (3) arginine-rich histones H3 and H4. This division may also be correlated with the degree of conservation of the histones' primary structure during evolution. On average, the histones are the most highly conserved of all proteins (Dayhoff, 1972; Wilson *et al.*, 1977). However, among the three groups mentioned, there are three degrees of conservation, with H1 being the most variable and H3 and H4 the most highly conserved.

1. Histone H1 is the most variable of the histones (for references, see Cole, 1977; Isenberg, 1978; Von Holt *et al.*, 1979), and the number of its subfractions varies from tissue to tissue and species to species. Three major domains can be distinguished within its primary structure: in the first domain, the amino-terminal region has charged residues and is variable; in the second domain, the central region has hydrophobic residues as in globular proteins and is not variable; and the third domain is the carboxy-terminal end which is charged and variable with respect to sequence conservation.

2. The slightly lysine-rich histones H2A and H2B are composed of two domains—the amino-terminal region, in which are clustered basic residues resulting in a net positive charge; and the carboxy-terminal region, which has a distribution of residues similar to other globular proteins and is rich in hydrophobic residues. This is demonstrated in Fig. 1 which gives the sequences of calf H2A (Yeoman *et al.*, 1972; Sautiere *et al.*, 1974) and H2B (Iwai *et al.*, 1972). H2A and H2B

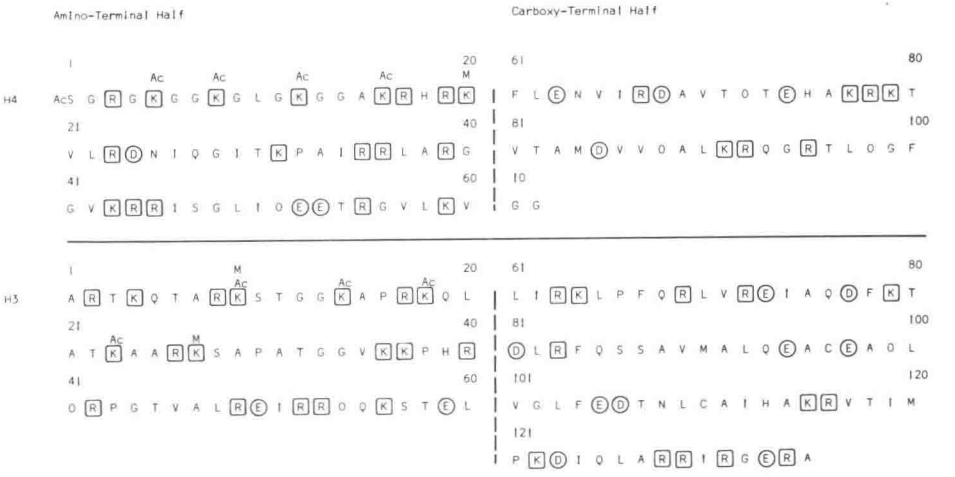

FIG. 2. The amino acid sequencies of calf histone H4 (DeLange *et al.*, 1972; Ogawa *et al.*, 1969) and H3 (DeLange *et al.*, 1972; Olson *et al.*, 1972; DeLange *et al.*, 1973). The notation is as in Fig. 1 except that superscript M denotes methylation site.

are conserved to an intermediate extent; there are variations between species. However, there are also subtypes of H2A and H2B which occur in the same species (Franklin and Zweidler, 1977) and whose synthesis is correlated in some species with different developmental stages (Cohen *et al.*, 1975; Newrock *et al.*, 1978). From the sequences known at present the carboxy-terminal region is the more conserved region (for references, see Isenberg, 1978; Von Holt *et al.*, 1979).

3. The arginine-rich histones H3 and H4 are the most highly conserved proteins during evolution (Patthy *et al.*, 1973). Like H2A and H2B, they are composed of two domains—the charged amino-terminal region and the hydrophobic carboxy-terminal region. This is shown in Fig. 2 which gives the sequences of calf H4 (DeLange *et al.*, 1969; Ogawa *et al.*, 1969) and H3 (DeLange *et al.*, 1972, 1973; Olson *et al.*, 1972).

II. Histone–Histone Interaction

A. Introduction

The chromatin repeating unit—the nucleosome core particle—is composed of an octameric unit of the four core histones (two copies each of H2A, H2B, H3, and H4) around which is folded 145 base pairs of DNA (for references, see Kornberg, 1977). This unit is rather compact, and one can *a priori* anticipate histone–histone interactions within the core of the nucleosome. The nucleosomes are further packed in the presence of histone H1 [which is probably bound on the outside of the nucleosome (Baldwin *et al.*, 1975)] to form hierarchies of higher order structures of chromatin. These higher order structures may involve interactions which are additional to those present in the nucleosome, i.e., histone–histone contacts between nucleosomes. A full description of the organization and conformational properties of the histone octamer awaits the three-dimensional solution of the structure of nucleosome core particle crystals (Finch *et al.*, 1977). Our understanding of histone organization and conformation within the higher order structures of chromatin is even more limited at present. Nevertheless, by probing with radiation and chemical cross-linking reagents, it has proved possible to analyze, albeit rather coarsely, the hierarchy of histone–histone contacts within and between nucleosomes. Some of these contacts can be preserved upon removal of the DNA. This permits the isolation of stable protein oligomers which are subcomplexes of the octamer. Similarly, the *in vitro* interaction of purified single histones gives rise to specific