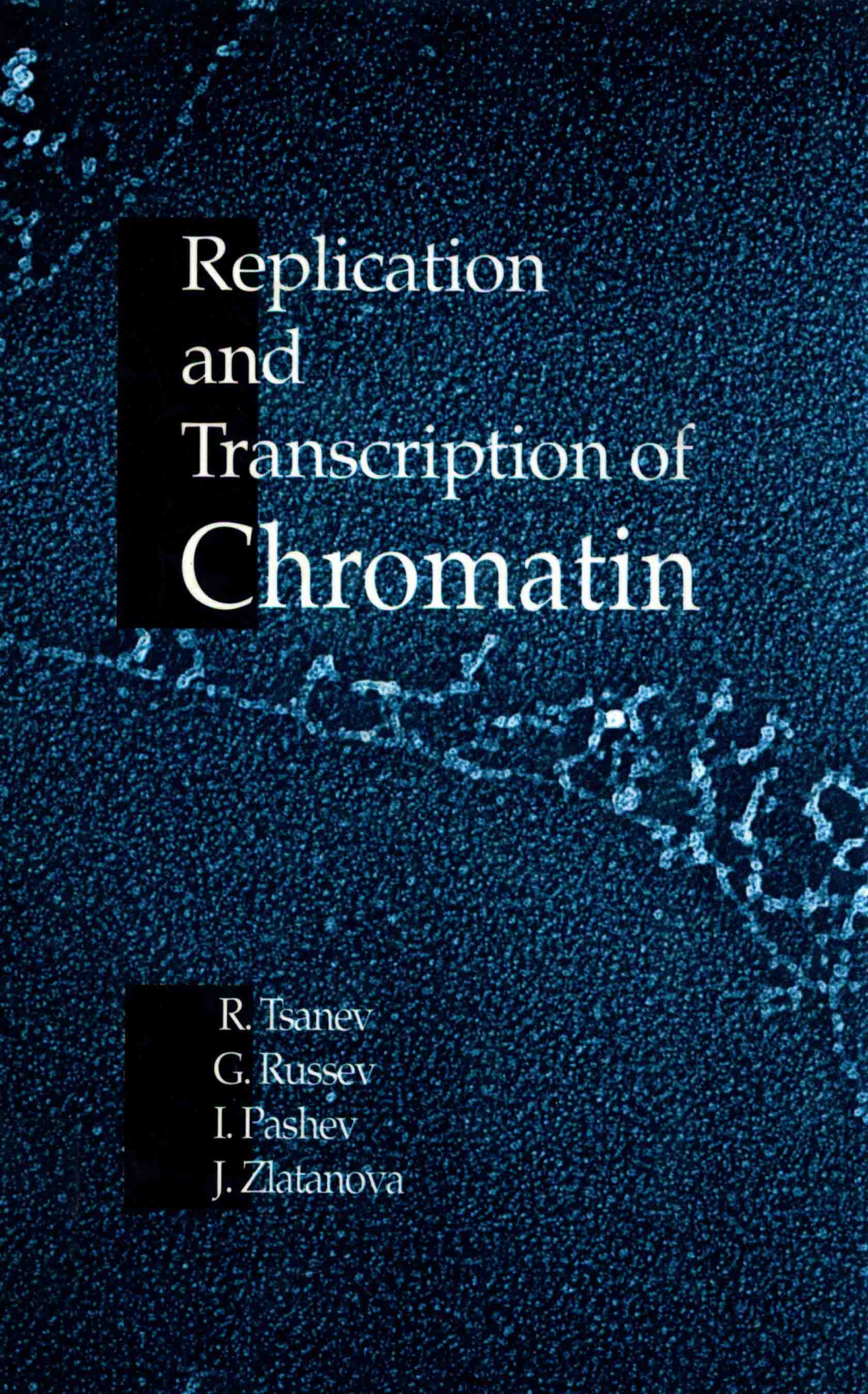
Replication
and
Transcription of
Chromatin
R. Tsanev
G. Russev
I. Pashev
J. Zlatanova

# Replication and Transcription of Chromatin

Roumen Tsanev, M.D., Ph.D., D.Sci.
Institute of Molecular Biology
George Russev, Ph.D., D.Sci.
Institute of Molecular Biology
Iliya Pashev, M.D., Ph.D., D.Sci.
Institute of Molecular Biology
Jordanka Zlatanova, Ph.D., D.Sci.
Institute of Genetics

Bulgarian Academy of Sciences
Sofia, Bulgaria


CRC Press
Boca Raton Ann Arbor London Tokyo

**Library of Congress Cataloging-in-Publication Data**

Replication and transcription of chromatin / Roumen Tsanev . . . [et al.].
p. cm.
Translated from the Russian.
Includes bibliographical references and index.
ISBN 0-8493-6803-0
1. Chromatin. 2. Chromosome replication. 3. Genetic transcription. I. Tsanev, Rumen.
QH599.R46 1992
574.87'322—dc20 92-17965
CIP



Direct all inquiries to CRC Press, Inc., 2000 Corporate Blvd., N. W., Boca Raton, Florida, 33431.



International Standard Book Number 0-8493-6803-0

Library of Congress Card Number 92-17965
Printed in the United States 1 2 3 4 5 6 7 8 9 0
Printed on acid-free paper

# PREFACE

During the last two decades, molecular biology has made exciting progress in unraveling the complicated structure of the genetic apparatus of eukaryotic cells. The interactions between the prime biological polymers — DNA, RNA, and proteins — have resulted in an hierarchy of structures which, on the one hand, have succeeded in accommodating a huge amount of DNA in a minute volume and, on the other hand, have created new problems for the functioning of the genome which were ingeniously solved by nature.

The efforts of scientists to reveal the two different aspects of that problem — to read the information encoded in DNA sequences and to understand how these sequences are integrated into a complex nucleoprotein structure — have resulted in many thousands of papers. Many excellent reviews have also appeared on selected topics. To add a new one may be justified by the rapid development in this field and by the need to look at some still disputable data from different viewpoints.

The aim of this review is to present the basic data concerning the molecular mechanisms that ensure the transfer of genetic information in eukaryotic cells, with an emphasis on the problems arising from the chromatin structure. It is also our intention to present not only facts but also conflicting results, hypotheses, and models proposed by different authors, as well as our own views on some problems with the belief that a new idea, even if later found to be wrong, may be as useful as new facts in stimulating thinking and new avenues of research.

The list of references cited in this book is necessarily restricted to a limited number of publications, mostly from the last twenty years. It by no means covers the whole literature, but all published data can be found in the papers quoted.

## THE AUTHORS

**Roumen Gueorguiev Tsanev, M.D., Ph.D., D.Sci.,** is Director of the Institute of Molecular Biology, Bulgarian Academy of Sciences, and Head of the Department of Cellular Regulatory Mechanisms.

Dr. Tsanev graduated from the Medical Faculty of the University of Sofia in 1947. He obtained training in Biochemistry in Budapest (1954), in Moscow (1957), and in Paris-Saclay (1963). He started his scientific work as a Research Fellow in the Institute of Biology, Bulgarian Academy of Sciences (1947 to 1954) and then he founded the Biochemical Research Laboratory that was later transformed into the Institute of Biochemistry (1972) and then the Institute of Molecular Biology (1978).

In 1964 Dr. Tsanev became a Professor of Biochemistry and has presented a number of invited lectures at international and national meetings. He has been an Oversea's Fellow of Churchill College in Cambridge (1969–1970), was elected a full member of the Bulgarian Academy of Sciences (1979), a foreign member of the Czechoslovak Academy of Sciences (1982), and a member of the London-based Academia Europaea (1989). He is also a member of the Council of ICRO (International Cell Research Organization at UNESCO), of ESGCP (European Study Group for Cell Proliferation), a reviewer for ICRETT (International Cancer Research Technology Transfer program of the International Union Against Cancer). He has been a member of the advisory boards of *Journal of Theoretical Biology* (1970–1980), of *European Journal of Biochemistry* (1972–1977), of *Cell Differentiation* (1978–1983) and of the Editorial Board of *Neoplasma*. He has received a state award for mathematical modeling in biology (1969) and a national award for achievements in biochemistry (1978).

He has published more than 180 scientific papers, primarily in international journals and books. His major scientific interests are the biochemistry of nucleic acids, the molecular organization of chromatin in relation to cell proliferation and differentiation, and the regulatory role of genetic networks.

**George Christov Russev, Ph.D., D.Sci.,** is Head of the Department of Molecular Biology of the Cell Cycle at the Institute of Molecular Biology, Bulgarian Academy of Sciences. Dr. Russev graduated in 1965 from the Department of Chemistry, University of Sofia. In 1971 he received his Ph.D. in biology from the Scientific Council of the Department of Biology of the Bulgarian Academy of Sciences and in 1982 he received his honorary D.Sci. degree.

Dr. Russev is currently Research Professor in the Institute of Molecular Biology, Sofia. He has received prestigious national and state awards in chemistry and biology and has been invited to give lectures at international meetings and at different universities in Europe, the United States, and Canada. His major scientific and research interests include chromatin structure, histone distribution during replication, and different aspects of DNA synthesis in eukaryotic cells. On these and related subjects he has published over 70 scientific papers.

**Iliya G. Pashev, M.D., Ph.D., D.Sci.,** is a Research Professor and Head of the Laboratory of Molecular Cytology in the Institute of Molecular Biology, Bulgarian Academy of Sciences, Sofia.

Dr. Pashev graduated in 1962 from the Higher Medical Institute in Plovdiv with his M.D. degree. He obtained his Ph.D. in biochemistry in 1968 from the same institution. In 1987 he received a D.Sci. degree from the Institute of Molecular Biology in Sofia.

Dr. Pashev has been an invited speaker at a number of international meetings and has presented many guest lectures at foreign universities and institutions. He has published more than 60 research papers. His major current research interests are focused on chromatin structure and function with an emphasis on the folding of DNA in the eukaryotic nucleus and the chromatin structure of transcribed genes.

**Jordanka Zlatanova, Ph.D., D.Sci.,** is founder and Head of the Molecular Genetics Department at the Institute of Genetics, Bulgarian Academy of Sciences, Sofia.

Dr. Zlatanova graduated in 1968 from Leningrad State University, USSR, Chair of Biochemistry. She obtained her Ph.D degree from the Institute of Molecular Biology, Bulgarian Academy of Sciences in 1980 and her D.Sci. degree from the Institute of Genetics, Bulgarian Academy of Sciences in 1989. In 1990 she was elected a full professor in molecular biology. Dr. Zlatanova is a member of the Bulgarian Union of Scientists, Biochemistry and Biophysics Section, of the Austrian Biochemical Society and of the International Society for Plant Molecular Biology. She was a recipient of an ICRETT (International Cancer Research Technology Transfer) Award.

Dr. Zlatanova has published more than 90 research and review papers and booklets. Her current major research interests include chromatin structure and function during cell proliferation and differentiation, and more specifically, the role of the lysine-rich histone H1 in regulating eukaryotic gene expression.

## ACKNOWLEDGMENTS

We would like to express our thanks to Professor K. E. van Holde for careful reading of some sections of the manuscript and for the useful comments on the part concerning nucleosome positioning. We would also like to thank Dr. Boyka Anachkova, Mrs. Eugenia Nikolova, Mrs. Antoaneta Kumpilova, and Mr. Z. Apostolov for their valuable help in preparing the manuscript.

# TABLE OF CONTENTS

Chapter 1

# MAIN FEATURES OF THE MOLECULAR ORGANIZATION OF CHROMATIN

## I. GENERAL OUTLINES

The transition from prokaryotes to eukaryotes has been associated with major changes in the organization of DNA. This was due to the interaction of the DNA with a specific type of proteins, which led to its much more complicated organization in the eukaryotic nucleus. These DNA-protein complexes, designated with the classical term *chromatin*, are no longer considered as a random association of DNA with proteins but as a regular structure with a stoichiometric DNA/protein ratio. It would be impossible to understand the mechanisms of DNA replication and transcription in the eukaryotic nucleus without knowing the molecular organization of these DNA/protein complexes.

The interaction of DNA with two different classes of proteins — histones and nonhistone proteins — leads to three levels of DNA folding: (1) nucleosomal level, (2) higher order structures, and (3) chromatin loops. This makes the path of DNA in the chromatin extremely complicated, raising difficult topological problems connected with the replication and transcription of the double helix. These problems are solved by a strictly ordered hierarchy of structures assembled by specific DNA-protein and protein-protein interactions, and by the activity of another group of proteins with specific enzymatic functions.

In this chapter, the main facts and conceptually important ideas concerning the molecular organization of chromatin will be only briefly reviewed as a necessary basis for understanding its functions and its problems (see reviews[1-4a]).

## II. FIRST LEVEL OF DNA FOLDING — THE NUCLEOSOME

### A. THE HISTONES

The evolutionary jump into the kingdom of eukaryotes has been associated with the appearance of a very specific class of basic proteins, the histones. It seems that the emergence of histones organizing DNA into specific subunits (nucleosomes) was a crucial step, which established a mechanism of DNA folding so advantageous that it has been preserved throughout the whole evolution of eukaryotes. Thus, it should be stressed that the problems of replication and transcription in eukaryotes are not problems of naked DNA but problems of a regularly organized DNA-histone complex. For this reason relatively more attention will be paid to histones.

#### *1. Main Characteristics of Histone Species*

There are five major histone species — the very lysine-rich group of histones H1/H5, the slightly lysine-rich histones H2A and H2B, and the arginine-rich

histones H3 and H4. Without going into details (see reviews[3,5,6]), several features of the histone structure are of interest for the structural organization of chromatin and its functioning.

Histones are basic proteins rich in lysine and arginine, poor in aromatic amino acids, and lacking tryptophan. Cysteine is present only in H3 at position 110 where it is evolutionarily conserved. A second cysteine residue has been found in calf thymus at position 96. As exceptions, H4 of echinoderms[7,8] and the H2B-like histone from testis[9] contain cysteine. A curious exception is the absence of cysteine in H3 of yeast,[10] which makes still more obscure the role of this amino acid.

Histones show a remarkable evolutionary stability, the most conservative protein being H4. While the mutational rate of most proteins is usually 30 to 400 point mutations per 100 amino acids per $10^9$ years,[11,12] this figure for H4 is only 1. With the possible exception of some lower eukaryotes[13] and the curious replacement of a threonine residue with cysteine in echinoderms, there are only two conservative mutations in H4 during the period separating plants and mammalians.[14]

Relatively large variations in their primary structure were found in the H2A-H2B pair. A comparison of the structure of H2B from evolutionarily different species ranging from molluscs to man (for references see[5]) shows that most of the substitutions have taken place in their tails, in the N-terminal domain of H2B and the C-terminal region of H2A, while the N-terminal tail of H2A and the C-terminus of H2B are well conserved in evolution even between yeast and man.[15,16] Nevertheless, the amino termini of H2A and H2B of yeast were shown to have interchangeable functions, probably because their positively charged domains shared similar structural features.[17] It should be stressed that, in these histones, the hydrophobic regions involved in protein-protein interactions are also relatively well conserved.

The very lysine-rich histones H1 are the most variable group, showing large variations in different species comparable to the variations of most other proteins. It should be pointed out that a microheterogeneity of all these histone species may not be real since it may be due to post-translational modifications of the molecule (e.g., compare [18] and [19]).

Thus, the evolutionary stability of different histone species obeys the general rule observed in globular proteins: the most conservative is the inner part of the molecule which is not in contact with the solvent; the most variable are the external sites.[20] Similarly, the order of evolutionary stability of histones — arginine-rich > slightly lysine-rich > very lysine-rich histones — is correlated with the localization of these histones in the nucleosome where the arginine-rich species form its inner core; the H2A-H2B pair is added laterally, while H1 is an external histone (see Section II.B.1). This may be explained by the need to conserve the hydrophobic sites involved in protein-protein interactions, assembling the nucleosome in a unique manner.

The basic amino acids are unevenly distributed along the polypeptide chain. The positive-charge density in all histones is higher at the ends of the

molecules — at the N-terminal part of the four nucleosomal histones and at the C-terminal part of H1.

At very low ionic strength, low pH, or in urea, histones have no ordered structure, while by increasing the ionic strength, ordered alpha structures are formed in the hydrophobic regions. NMR studies[21-25] and tryptic digestion (see review[26]) have shown that all core histones have unstructured N-terminal tails in a random-coil conformation, which are able to interact with DNA, and structured globular regions in the C-terminal part. H2A and H3 have also short C-terminal tails. Three distinct structural domains are formed in H1: an N-terminal random-coil domain ("nose"), a globular part in the center ("head"), and a random-coil C-terminal end ("tail").

Thus, an important structural feature of histones is that two main distinct domains are formed in all histone molecules — charged random-coil ends with potential sites for protein-DNA interactions and ordered hydrophobic alpha structures involved in protein-protein interactions. These interactions are stronger and more specific between H3 and H4, so that these histones easily form extremely stable tetramers in high salt solutions. H2A and H2B also form specific dimers which are less tightly held between themselves and to the H3/H4 tetramer.

The histones are subject to posttranslational modifications of different amino acid residues, mainly $N^6$-acetylation and $N^6$-methylation of lysine, O-phosphorylation of serine and poly-ADP-ribosylation (see review[27] and Chapter 3, Section VI). These modifications lead to neutralization, introduction or retention of charge and are supposed to play some functional roles (see Chapter 3, Section VI). Compared with the strictly conserved positions of the charged residues in the histone molecule, these changes in charge were considered to be paradoxical.[28] However, this fact should not be surprising. It only supports the view that the functions of chromatin would require regular charge changes in the nucleosome at strictly fixed sites involved in DNA-protein interactions.

With the exception of H4, all other histones are represented by a few nonallelic histone variants (isohistones) detected by using gel electrophoresis in the presence of Triton X100.[29-34] In most cases, H2A is represented by two and H3 and H2B by three variants, while H4 has no variants (Figure 1). It is important to point out that the isohistones differ by several aminoacid substitutions, mainly in their hydrophobic regions, thus affecting the sites of protein/protein interactions, which are evolutionarily the most stable sites. This may be an important event influencing the possibility of conformational changes of the nucleosomes necessary for gene functioning (see Chapter 2, Section IV.B and Chapter 3, Section V.A.3).

A characteristic feature of histones is that (with the exception of yeast) they are encoded by multiple copies of their genes (see Chapter 2, Section IV.A.2).

### 2. *Occurrence of Histones*

The first living organisms with typical histones in their nuclei are the lower eukaryotes. In all classes studied — algae, fungi, slime molds, and protozoa —

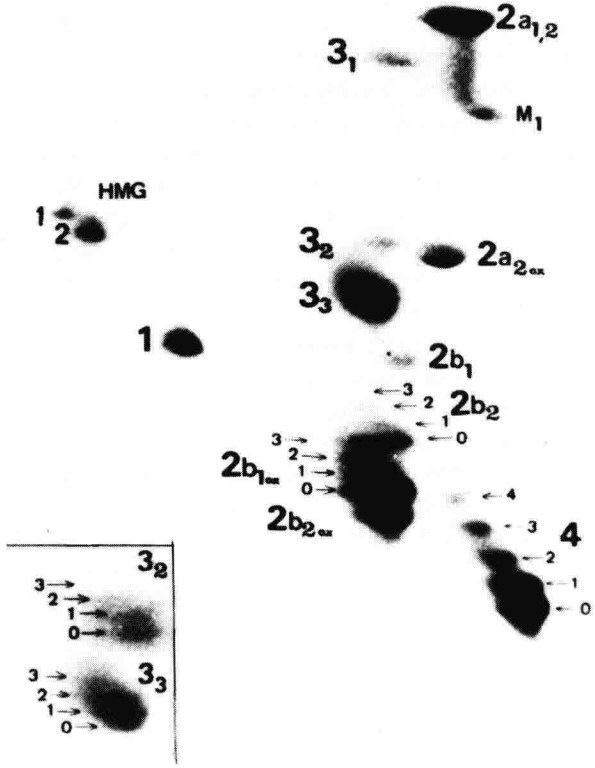


**FIGURE 1.** Two-dimensional electrophoretic separation of histones, showing the presence of histone variants (two variants of H2A — $2a_1$ and $2a_2$; three variants of H3 — $3_1$, $3_2$, and $3_3$; two variants of H2B — $2b_1$ and $2b_2$) and their acetylated forms (indicated by arrows, indicating also the number of acetyl groups) in the chromatin of a putatively active chromatin fraction from EAT cells. First dimension (left to right) in 15% polyacrylamide gel in acetic acid/urea; second dimension (top to bottom) in the presence of Triton X100. (From Georgieva, E. I., Pashev, I. G., and Tsanev, R. G., *Arch. Biochem. Biophys.*, 216, 88, 1982. With permission.)

a full complement of the main histone species has been well documented, although in these species the histones show more structural variations as compared to the histones of higher eukaryotes (see reviews[5,35]).

The only interesting and curious exception is the *Dinoflagellates*, where no histones have been found. These organisms were considered to be mesokaryotes, an intermediate form between prokaryotes and eukaryotes. However, new data on their ribosomal RNA have indicated a close relationship between *Dinoflagellates*, *Cilliates* and yeast.[36] This fact supports the suggestion that *Dinoflagellates* are, in fact, lower eukaryotes which have lost the whole cluster of their histone genes. If strictly proved, this will be a striking example that histones may be dispensable in the evolution of lower eukaryotes.

A controversy exists concerning the presence of histone H1 in yeast, where in most studies it has not been found. Some authors have reported histone-like

proteins in this unicellular eukaryote, but with conflicting data on their properties (see references in[38]). A more recent study has demonstrated a protein in yeast which, according to its electrophoretic mobility and immunological reactivity, seems to be an H1-like protein.[38] This protein is very easily degraded by proteolytic enzymes, which explains the failure of its detection.

In all higher eukaryotes, both in plants and in animals, the main histone species are present with no exceptions. A marked difference between plants and animals exists in their H2A-H2B histones, which differ in electrophoretic mobility and show some changes in their primary structure.[39-41] The plant H2B was found to be very similar to the corresponding calf thymus histone in its middle and C-terminal hydrophobic regions, the N-terminal part being different.[42] The same was suggested for H2A. In spite of these differences, the plant histones are structurally and functionally equivalent to the animal histones — the plant chromatin has the same core nucleosome, with the standard DNA length of 145 bp,[43,44] as the animal chromatin. An interesting finding is the presence of a duplicate H2A-H2B pair in the slime mold *Dictyostelium discoideum*. One of these pairs migrates in urea/polyacrylamide gel as plant histones, the other as animal histones.[45] These data have suggested that this lower eukaryote may represent the species where, in evolution, the divergence between plants and animals took place.

### 3. *Tissue Specificity of Histones*

No tissue-specificity exists concerning the four main histone species, H2A, H2B, H3, and H4, both in normal and in malignant tissues. Tissue differences exist with respect to the relative proportions of the nonallelic histone variants, whose significance is still unclear (see Chapter 3, Section V.A.5).

In contrast to the species differences of histone H1, it seems that this histone has the same subfractions in different tissues. The reported differences in some cases seem to be due either to incomplete extraction and proteolytic degradation of some subspecies or to chemical modifications and differences between proliferating and resting cells. The only well-documented case of a tissue-specific histone of this group is the histone H5, which appears during the terminal differentiation of nucleate erythrocytes. Still, the possibility of some tissue-specific differences in the pattern of H1 subfractions should not be neglected.

The only exception is the reproductive tissue, where drastic changes take place in the basic nuclear proteins of the male gametes during spermiogenesis (see references in review[5]). In all species studied, both in plants and in animals, the male gametes contain different sets of sperm-specific basic proteins: either sperm-specific histones (e.g., in echinoderms and in plant microsporocytes), histone-like proteins (e.g., in molluscs), or other basic proteins, like the protamines of fish and the cysteine-containing mammalian protamines. It is interesting that even among related species there are no two species with identical sperm-specific basic proteins, which in all cases serve the role of functionally inactivating, packaging, and protecting DNA in the sperm.

The primary structure of the sperm H2B histone of echinoderms[47] and the sperm proteins of molluscs[48] support the idea that the fish protamines, and

probably other sperm-specific proteins, have evolved from the repeated basic pentapeptide at the N-terminal region of the echinoderm H2B by duplications and point mutations.

### *4. Origin and Evolution of Histones*

After more than four decades of studies on these proteins, it is still a mystery how they appeared in evolution. DNA-protein interactions and complexing occurred previously in prokaryotic organisms as one of the fundamental processes controlling the function of the genetic apparatus — the interaction between *trans*-acting regulatory proteins and specific DNA *cis* sequences.

In addition to these proteins, another group of DNA binding proteins in prokaryotes has been described. These are often called histone-like proteins due to their basic character and small molecular mass (see reviews[5,49,50]). One of the prokaryotic DNA-associated proteins, HU, was shown to be able to organize DNA in nucleosome-like structures,[51] but it is not clear whether such structures exist *in vivo*. In some of its properties, the low-molecular weight HU resembles H2B and was found to induce negative superhelical turns in DNA, condensing it in a ratio similar to that caused by the four eukaryotic histones.[51] Some of these proteins were able to stabilize DNA against thermal denaturation.

All of these prokaryotic basic DNA-bound proteins seem far from being possible precursors of the eukaryotic histones. Evidently they have satisfied the needs of a unicellular prokaryote with about a thousand-fold less amount of DNA and the absence of the complex eukaryotic program of development and morphogenesis. The same seems to hold true for the basic DNA-bound proteins of *Dinoflagellates*, in view of the possibility that they may have lost their histone genes (see Section II.A.2).

It is very unlikely that each histone species has evolved separately from a different precursor. This observation is based on the unique and evolutionarily conserved structure — the nucleosome — assembling together four histone species with a piece of DNA (see Section II.B.1) and the strong selection pressure against histone mutations, in spite of the presence of multiple copies of histone genes[52] (see Chapter 2, Section IV.A.2). The lack of mutations in the several hundred copies of histone genes is very suggestive: it shows that when DNA is organized in nucleosomes, the presence of even one "false" histone molecule would produce a structure preventing the function of the genome. These data indicate that the full deletion of the histone genes (as in the case of *Dinoflagellates*, see Section II.A.2) is compatible with a primitive life, but the presence even of a small number of incorrect histone molecules is lethal.

Thus, it is highly improbable that several independent lines of mutational changes may have led to a structure which does not permit deviations from a unique design. It seems much more probable that the four nucleosomal histone species have evolved from one or two common precursors. The finding of homologies between the major histone species and the specific clustering of their genes[52] (see Chapter 2, Section IV.A.2) has already suggested that they may have evolved from one or two common precursors.[28,53,54] It could be speculated