Cell Biology Monographs

Volume II

J. Herbert Taylor

DNA Methylation and Cellular Differentiation

11

Springer-Verlag
Wien New York

DNA Methylation and Cellular Differentiation

J. Herbert Taylor

Springer-Verlag
Wien New York

Prof. Dr. JAMES HERBERT TAYLOR
Institute of Molecular Biophysics,
The Florida State University, Tallahassee, Florida, U.S.A.

This work is subject to copyright.
All rights are reserved, whether the whole or part of the material is concerned,
specifically those of translation, reprinting, re-use of illustrations, broadcasting,
reproduction by photocopying machine or similar means, and storage in data banks.
© 1984 by Springer-Verlag/Wien
Printed in Austria by Adolf Holzhausens Nfg., Wien

With 17 Figures

Library of Congress Cataloging in Publication Data. Taylor, J. Herbert (James Herbert), 1916— . DNA methylation and cellular differentiation. (Cell biology monographs, ISSN 0172-4665; v. 11.) Bibliography: p. Includes index. 1. Deoxyribonucleic acid— Metabolism. 2. Methylation. 3. Cell differentiation. I. Title. II. Series. QP 624.T 39. 1984. 574.87'3282. 83-20239

ISSN 0172-4665

ISBN 3-211-81761-1 Springer-Verlag Wien — New York
ISBN 0-387-81761-1 Springer-Verlag New York — Wien

Preface

In 1977 I wrote a grant proposal in which I applied to study developmental patterns in enzymatic methylation of DNA in eukaryotes. One part of the proposal was to assay cells at different embryonic developmental stages for maintenance and *de novo* type methylase activity. With one exception the referees, probably developmental biologists, recommended that the work not be supported because there was no evidence that methylation plays any role in eukaryotic gene regulation. Aside from proving that innovative ideas can seldom be used to successfully compete for grant funds, the skepticism of biologists toward methylation as a regulatory mechanism was, and still is, widespread even among some of those who investigate the problem. That is a healthy situation for all points of view should be brought to bear on a problem of such importance. However, to deny funds to investigate a problem because one has already formed an opinion without evidence is hardly commendable. The great skepticism about the significance of DNA methylation is based in part on the evidence that it is absent or very little used in *Drosophila,* a favorite organism for genetic and developmental studies. There now remains little doubt that methylation of cytosine in certain CpG sites can strikingly affect the transcription of sequences 3′ to the methylated doublet. How this inhibition operates and to what extent it is utilized in cells is still debatable. Furthermore, a mechanism for the inheritance of a pattern of methylation once it is installed is understood and demonstrated in principle. What other functions DNA methylation may have is debatable. Certainly most methylation of CpG sites does not appear to affect transcription and many plants, which have up to one-third of the cytosines methylated, provide an enigma.

This little book is not a review of everything published on DNA methylation. Several short reviews have been published in the last several years (Taylor 1979, Razin and Riggs 1980, Yuan 1981, Razin and Friedman 1981, Doerfler 1981, Adams and Burdon 1982, and Riggs and Jones 1983). The interested reader may consult these reviews for further details. What I have tried to do is to survey the literature and to pick those investigations which are relevant to cellular differentiation, or provide background for understanding the problems. I have then picked a few examples and tried to relate them to a differentiating system and to see how the two fit together. Investigations are few and much remains to be learned, I am sure. Probably the most frustrating aspect of studies of DNA methylation is the paucity of methods for investigating the 5-methyl cytosine in genes before cloning. The methylated cytosine is lost when sequences are cloned and amplified for sequencing, because bacterial hosts for the cloning vectors do not have the maintenance enzymes to reproducibly methylate all or a major fraction of the CpG doublets in DNA. Perhaps that situation points to our best hope for the future. If the maintenance methylases can be characterized, isolated and cloned, we may be able to develop hosts for cloning vectors with the eukaryotic maintenance methylases. However, we are far from that stage today. We must be satisfied with a few restriction enzymes that are sensitive to methylation at CpG or CpC sites and to

the sequencing of a few DNA repeated sequences that can be prepared in enough copies for sequencing without amplification. Another recent development with great promise is the sequencing of single copy genes. It involves an adaptation of the Southern blot of a gel after a Maxam-Gilbert type reaction in which the probe is a fragment at one end of the sequence to be determined.

For those who read prefaces and do not like to read books from the end to the beginning, let me state my conclusions very briefly. I have concluded that DNA methylation is an important mechanism of cellular differentiation in vertebrates, although it probably is not in insects such as *Drosophila*. Even in vertebrates, methylation is only a subsidiary mechanism. The primary mechanism is based on a very simple scheme that eukaryotic cells invented hundreds of millions of years ago and cell biologists and geneticists have studied for years, most of them without understanding its real significance. Cells have partitioned their DNA into two pools for replication and packaging, an early replicating pool and a late replicating pool with a short pause between. It is possible that some cells have invented a third or fourth pool, although I doubt that elaboration is either necessary or desirable. In the first pool, replicated in S_E, are all of the genes that the cell will need to use in its differentiating processes and functional roles. All other genes replicate in S_L, the last half of the S phase, and are maintained in an inactive state. The proteins that sequester genes from transcription are available in S_L. The proteins that open up the genes to transcription are available only in S_E and any sequence replicated in S_E is potentially functional when and if the cellular environment is appropriate. This situation makes the control of the time of replication in the cell cycle crucial, but it probably makes little if any difference in what part of S_E or S_L a particular gene replicates. It also means that a mechanism to modify replication origins in a way that is stably inherited is a crucial feature of differentiation. Some replicons must be shifted to replicate in S_E, others to S_L.

Methylation as a part of differentiation may be a late development in evolution and is only extensively exploited in vertebrates in which it suppresses those genes in a cluster (replicon) that will not be useful in a particular differentiated cell. For example, in a red cell precursor all globin genes will be switched to replicate in S_E, but only those will be expressed which are demethylated during a subsequent determination step.

I know that both of these ideas, and particularly the first, will be considered naive and unsupported by all of the evidence by some critics, but I predict that biologists will still be investigating both these phenomena and acquiring astonishing results when the critics and I have passed from the scene.

I thank those who have cooperated with me by sending illustrations; each is credited in the legends of the illustrations except where there are multiple authors and it may not be obvious which one helped. I would also like to give credit to Dr. M. Busslinger and Dr. Gerhart Ryffel, who provided me with valuable information prior to its publication, and to Mrs. Myrna Hurst, who learned to use a word processor to record and revise the manuscript. I wish to acknowledge a one-half year sabbatical leave from Florida State University that enabled me to devote most of my time to the writing of the monograph.

Tallahassee, Florida, October 1983 J. Herbert Taylor

此为试读,需要完整PDF请访问: www.ertongbook.com

Contents

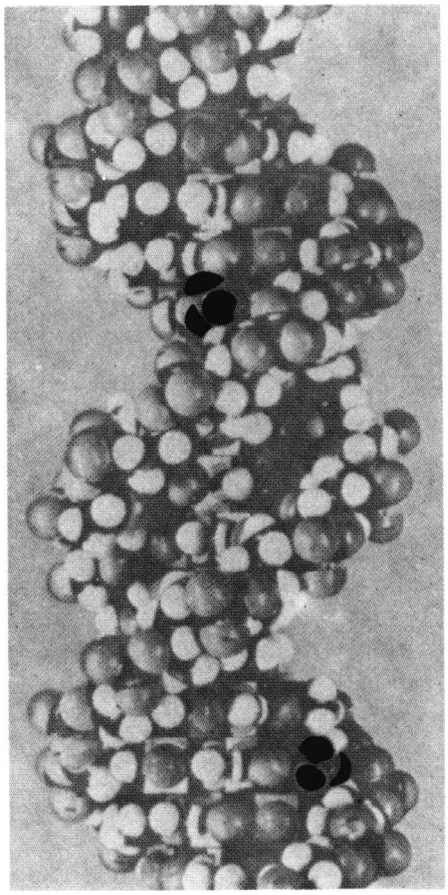

Fig. 1. Methylated DNA. Location of the methyl group of 5-methyl-2′-deoxycytidine in the B-form of a DNA helix. The methyl group projects into the major groove and occupies a position analogous to the methyl group of thymidine. Therefore, the conversion of deoxycytidine in DNA to 5-methyl-deoxycytidine makes the G : C base pair resemble an A : T base pair and a molecule binding to DNA at that point should encounter the difference. (From RAZIN and RIGGS; copyright 1980 by the American Association for the Advancement of Science)

I. DNA Methylation and Cell Differentiation: An Overview

A. Introduction

The problem of cellular differentiation has renewed interest for molecular biologists who are studying the modifications of DNA. Three modifications are getting much of the attention. One is a structural modification of the one dimensional message by the insertion of transposable genetic elements. Another is a base modification by methylation which does not alter the one-dimensional coding properties, but may change the affinity for other molecules and thereby transcription or readout of the code. The third is a new three dimensional structural form called Z-DNA which is related to the base sequence, the modification by methylation, the torsional properties and the molecular environment. This change would affect affinity for other molecules and could also affect the readout of the code.

We will be concerned here primarily with modification by methylation, but we can not ignore the other modifications and hope to understand the roles of methylation. All three modifications are reversible, but the insertion of transposable genetic elements requires a precise removal of the insert to restore the original sequence and these appear to be rare events. The modification by methylation is probably removed by dilution during semiconservative replication in the absence of methylase activity at certain sites. The conversion from the Z form back to the more usual B form requires a change in the molecular environment, but could be influenced by methylation and the torsional state of the chains. It is unlikely to persist through replication, but could reform afterwards if the conditions were favorable.

The modification of major concern to us is the methylation of cytosine at the fifth carbon in the DNA polymer after replication. In the duplex these methyl groups project into the major groove of the double helix (Fig. 1) and make the cytosine of a C:G base pair have a conformation more like a thymine residue of a T:A base pair than the usual cytosine. The maintenance of a methylation pattern is assumed to operate by an enzyme that inserts the methyl group on a new chain symmetrically to the one

retained on the parental chain. A highly specific enzyme could make the replacement almost as efficient as replication of a base sequence and the maintenance of patterns observed so far suggests a high fidelity. At some time when a new pattern is initiated, the enzyme would presumably require highly accurate sequence recognition to avoid complete methylation of all cytosines, which of course does not occur. Since most cells are unlikely to have demethylating enzymes, it is generally assumed that loss or change of a pattern occurs by a failure to maintain the pattern at specific sites when replication is occurring. After two rounds of replication, only two of the four segments of the DNA helix will have the hemimethylated DNA with a pattern for an enzyme to act upon. The pattern would be lost permanently from the other two segments (see section C, below). Here also sequence specificity would have to play a role in the positioning of a site specific inhibitor of the enzyme, which normally restores the modification on the new chains after replication, since the methylase activity must always be present for methylation of other sites after replication.

B. What New Properties Does Methylation Confer on DNA?

In 1948 HOTCHKISS (HOTCHKISS 1948) discovered that calf thymus DNA contained, in addition to the four principal purine and pyrimidine bases, a small amount of 5-methylcytosine (Fig. 2 shows a nucleotide and related molecular), and a few years later WYATT (1951) found higher amounts in DNA from wheat germ. In 1955 DUNN and SMITH reported another minor base, N6-methyladenine (N^6A), in bacterial DNA. When the enzymatic mechanisms of DNA replication were being investigated by KORNBERG and associates (BESSMAN *et al.* 1958), several analogs of the four principal bases were prepared as precursors for tests of incorporation in their in vitro system. Among these was 5-methyl-2′-deoxycytosine triphosphate. The nucleotide was incorporated into DNA in the in vitro system in place of cytosine at rates comparable to the four principal bases. Both of these modified bases form base pairs with little effect on the double helix and both are nearly as stable as the four principal bases in DNA. 5-Methylcytosine (5 mC) substitutes completely for cytosine in the *Xanthomonas* phage, Xp12 (Kuo *et al.* 1968). Before incorporation into the phage DNA the deoxycytidylate of the host cell is converted to 5-methyl-deoxycytidylate by an enzyme designated deoxycytidylate methyl transferase. The methyl group is derived from tetrahydrofolate which is converted to dihydrofolate in the reaction (FENG *et al.* 1978). The phage nucleotide containing 5mC is synthesized in a manner that resembles the synthesis of thymidylate from deoxyuridylate by the ubiquitous enzyme, thymidylate synthetase. The DNA with 5mC substituted for cytosine has a lower buoyant density in CsCl than predicted from other DNA with the same percentage of adenine plus thymine, and the thermal stability is increased. The melting

此为试读,需要完整PDF请访问: www.ertongbook.com

temperature is reported to be 83.2 degrees C in 0.012 M Na^+, which is the highest for any naturally occurring DNA and 6.1 degrees higher than regular DNA with the same percentage of adenine and thymine (EHRLICH *et al.*1975). The small amount of 5mC (about 0.5 mole percent) in most DNA from animal cells would be expected to have very little effect on these properties, but in some higher plants where the average 5mC content can be as high as 6–7 mole percent, *i.e.*, every third cytosine modified, the effect could be significant. Considering the probability that the 5mC is not uniformly distributed the effect on the physical properties could be

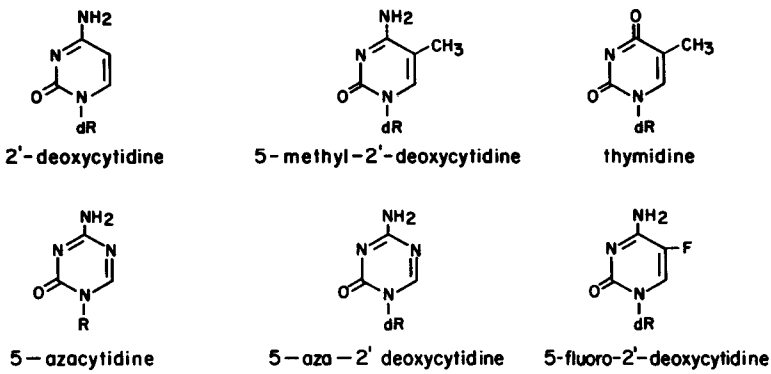

Fig. 2. The tree pyrimidine nucleosides found in DNA and three analogs which can inhibit methylation of cytosine in DNA. 5-azacytidine is not found in DNA, but 5-aza-2'-deoxycytidine diphosphate can be derived from the enzymatic reduction of 5-azacytidine diphosphate and the triphosphate is a precursor for the replacement of 2'-deoxycytidine or 5-methyl-2'-deoxycytidine. 5-fluoro-2'-deoxycytidine can also replace the same nucleosides and the enzymatic methylation should be blocked at the 5 position by the fluorine which replaces hydrogen

dramatic. In addition, BEHE *et al.* (1982) reported that methylation of the synthetic polynucleotide poly (dG:dC)·poly dG:dC) can have a striking effect on its transition from the B form to the Z form. The unmethylated polymer can be converted to the Z form only in a high salt concentration, but the methylated polymer can be converted at a concentration approaching the expected physiological conditions in cells. The polymers used in the experiments were synthesized from nucleotides containing the methylated base. However, the methylated DNA from cells, with the exception of that from the phage Xp 12 mentioned above, is modified by methylation of the polymer after replication by one or more of the class of enzymes called methyl transferases or what we shall usually refer to as DNA methylase.

MANDEL and BOREK (1963) described an enzymatic activity which transferred the methyl group from S-adenosylmethionine (SAM) to

polymeric RNA. This discovery initiated a number of studies on the in vitro methylation of transfer RNAs and other types of RNA which is outside the area of our discussion. The methylation of DNA was demonstrated in vitro by GOLD *et al.* (1963 a, b) and by GOLD and HURWITZ (1963) in extracts from bacteria. The donor of the methyl group as for the RNA methyl transferase was SAM. After the modification-restriction system of some phages which infect *E. coli* was shown to be due to methylation of the DNA (LINN and ARBER 1968 and MESELSON and YUAN 1968), the significance of the methyl group for controlling the binding of proteins to DNA became apparent. The methylase, now called type I, binds to DNA which is unmethylated at specific recognition sites, but the reaction is a slow one. However, once bound the enzyme is not released and depending on intracellular conditions may either methylate the DNA or restrict (cleave) it. On the other hand if the DNA is hemimethylated (methylated on one strand at the recognition site), the enzyme rapidly methylates the other chain, but will not restrict (YUAN 1981). The demonstration a few years later that type II restriction endonucleases were also inhibited by methylation of either adenine or cytosine at specific sites which varied for each restriction enzyme (SMITH and WILCOX 1970, and KELLY and SMITH 1970) further emphasized the role of the methyl group on these two bases in regulating the binding as well as the catalytic activity of specific proteins at specific sites on double-stranded DNA. Models of the double helix show that the methyl groups of both 5mC and N^6A project into the large groove (Fig. 1). In the case of 5mC the methyl group at a binding site probably resembles that of thymine as shown by the experiment described below.

An experiment on the binding of *lac* repressor to its specific site on the *E. coli* DNA illustrates this change in affinity in a dramatic way. FISHER and CARUTHERS (1979) studied the affinity of *lac* repressor protein which functions in the regulation of the lactose operon in *E. coli*. Of the 26 nucleotide pairs in the *E. coli* chromosome which constitute the binding site for the *lac* repressor, the A:T pair at position 13 (Fig. 3) is crucial. Mutations with a substitution of a G:C base pair at that position are constitutive for the enzymes coded by the *lac* operon. The binding is insufficient to block expression of the genes even in the absence of lactose. In vitro experiments in which a uracil residue was substituted for the thymine showed that the affinity for the *lac* repressor was dramatically reduced. A tenfold reduction in the stability of the repressor-operator complex was observed. The repressor-operator interaction was assumed to be hydrophobic and consistent with that idea they found that bromouracil substituted for thymine produced a slight reduction in the stability of the complex. In a transversion, in which the adenine-thymine pair was changed to thymine-adenine, there was a seven fold reduction in the stability of the complex. This suggests that the hydrophobic interaction is positionally specific. When 5mC was substituted for cytosine in position 13, the stability

of the complex was equal to that of the standard wild type with the A:T base pair at that position. These reconstructions of the site demonstrate that the conversion of a G:C base pair to G:5mC by methylation can have a functional effect similar to a mutation in which the base pair G:C is converted to A:T. However, by the dilution of the methylated base in replication without the action of a methylase, the mutant phenotype would return. We may conclude that the methyl group projecting into the large groove of DNA can make a critical difference in the binding of various specific proteins to double-stranded DNA.

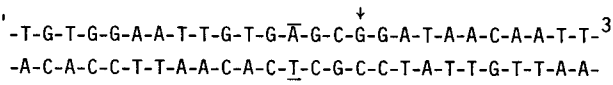

Fig. 3. The *lac* operator DNA sequence from the *E. coli* at which the *lac* repressor protein binds. The arrow indicates the axis of the twofold symmetry in the region and the line above and below the sequence indicates a base pair which when changed by mutation from A:T to G:C has a marked effect on binding of the repressor. (Adapted from FISHER and CARUTHERS, 1979)

C. The Origin and Maintenance of Methyl Cytosine in DNA

Since the methylcytosine content of DNA varies among species and among different fractions of DNA in the same genome, it could influence the types of proteins which interact with the DNA. The effects of the proteins on the activities of the DNA could in turn change the phenotype. Therefore, we shall examine briefly the ways in which DNA can be modified and how such modifications could be inherited from cell generation to cell generation. An important consideration will also be the stability of the pattern once it is established. All DNA appears to be synthesized with deoxycytidine triphosphate as a precursor of the deoxycytidine nucleotides, with the exception of DNA in a few bacterial phages. Therefore, the pattern of methylation must be imposed by enzymes that act post-replicatively. As first indicated by the catalytic behaviour of the Type I methylating enzymes of bacteria, there is a basis for the hypothesis that hemimethylated DNA might react with methylases in a unique way. Early experiments indicated that cytosine was the only site of enzymatic, post-replicative methylation in the DNA of most eukaryotes. In animals more than 90% of the 5mC found in DNA is in the dinucleotide 5' pCpG (DOSKOCIL and SORM 1962, GRIPPO *et al.* 1968), which we will refer to as the CpG doublet. If the DNA is methylated on both chains at CpG doublets, the two new helices after replication will be hemimethylated as shown in Fig. 4. An enzyme, with a preferential affinity for the hemimethylated site and an efficient catalytic activity for adding a methyl group to the symmetrical cytosine, could

maintain the pattern with a high fidelity from one cell generation to the next. We will designate such enzymes "maintenance enzymes" (HOLLIDAY and PUGH 1975, and TAYLOR 1979). One of the crucial questions will be, do such enzymes exist in replicating cells and if so how faithfully are the patterns maintained?

If the control of DNA function is related to the pattern of 5mC in eukaryotes, there must be enzymes which initiate the patterns. In bacteria all of the methylases so far known can function to initiate a pattern by reacting

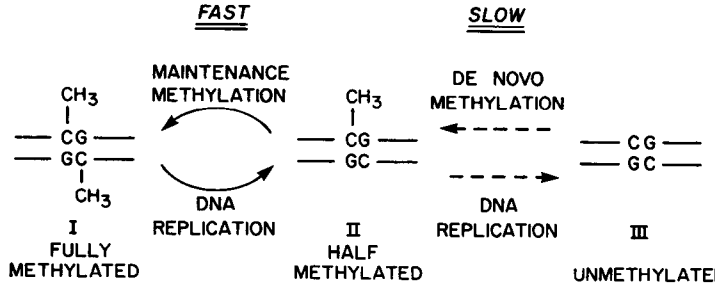

Fig. 4. Three possible states of DNA methylation: (I) fully methylated, (II) half methylated or hemimethylated and (III) unmethylated. Fully methylated DNA would be converted to hemimethylated DNA by ordinary replication which does not utilize 5-methylcytosine containing nucleotides as precursors. However, if a maintenance methylase is available, the hemimethylated DNA will be converted to a fully methylated form before the next replication. If for any reason this does not happen, both chains will be unmethylated after another replication. Maintenance methylase can not restore a fully methylated pattern unless one strand is already methylated. (From RAZIN and RIGGS; copyright 1980 by American Association for the Advancement of Science)

with specific recognition sites, many of which are now known. There is usually a corresponding endonuclease which recognizes the same sequence and cleaves the DNA unless it is methylated on at least one chain. Such endonucleases, which are called restriction enzymes, are rare if they occur at all in eukaryotes. Furthermore, the methylatable sites for a particular enzyme are very efficiently methylated in prokaryotes (bacteria), while in eukaryotes the sites are incompletely methylated. However, methylation patterns appear to be consistent for a particular type of cell at a particular stage in development or differentiation as will be illustrated later on in this monograph.

Methylases which initiate new patterns of methylation have been referred to as de novo (RAZIN and FRIEDMAN 1981) or initiation type methylases (TAYLOR 1979). In prokaryotes, both functions, initiation and maintenance, may be carried out by the same enzyme complex. Not enough is yet known

about the methylases of eukaryotic cells to decide this question. However, there is considerable evidence that maintenance of a pattern does occur once it is established and there is circumstantial evidence that some patterns have to be initiated at certain stages in development. Research on the properties of eukaryotic methylases is still in a primitive state in spite of a few interesting and provocative reports (GRUENBAUM *et al.* 1982). Fig. 4 shows schematically how initiation, maintenance and deletion of a pattern might occur.

The other question that will require an answer is how methylation patterns at specific sites within a gene can be changed while similar sites in an adjacent gene are maintained. Since there is only one unconfirmed report of a demethylating enzyme which acts on polymeric DNA (GJERSET and MARTIN 1982); it seems too early to conclude that such enzymes operate to eliminate methylation at specific sequences. It is more likely that methylation is lost by dilution during replication, but the question of a mechanism which would allow specificity for such a dilution remains unanswered. One possibility is a site specific inhibitor, but would cells code enough inhibitors for each individual gene that must be changed to an unmethylated state? Since that seems unlikely, a smaller number of inhibitors, that bind to families of similar sequences and inhibit methylation downstream to the next binding sequence while replication proceeds through two rounds would effectively eliminate the pattern from both strands on two of the four daughter chromosomes or more specifically on certain sequences of those chromosomes (Fig. 4). Available evidence suggests that changes in methylation patterns, either the loss or gain, are associated with cellular proliferation and therefore, with DNA replication.

D. Differentiation: The Problem Posed

The problem has always been to explain how a single cell, the zygote, with a single genotype can produce a highly organized arrangement of cells with the many phenotypes that make up the body of the individual eukaryotic organism. As long as biologists have accepted the cell theory and begun to understand how cells reproduce and segregate into functional organs, the origin of this diversity has puzzled and intrigued them. The first notably philosophical solution was suggested by WEISMANN (1892) in the last century when he proposed that the genetic determinants were segregated to the different cells in some systematic fashion. With the advent of the chromosome theory of inheritance and the concept of the genes which make up a chromosome, there was a physical basis for this concept. However, as the knowledge of chromosomes developed, the evidence indicated that with few exceptions the chromosomes remained morphologically the same in all of the cells of an individual. There were some notable exceptions such as the discovery that the germline and somatic cells had a different chromosome

complement in a few cases. For example, during the early cleavage stages in the embryo of the parasitic worm, *Ascaris,* there is an elimination of parts of the chromosomes in cells that will form the somatic organs. However, the full complement is retained and passed on in the cells of the germline. In some insects, notably certain species of dipterans, there is an elimination of whole chromosomes from the cells that will differentiate into somatic tissues. Of course, this could only account for a difference between the germline and the somatic cells, but polyploidy and polyteny were also discovered in a wider variety of cells. Considerable discussion of the possible significance of these changes can be found in the literature before 1950. The multiplication of all of the genes could not give rise to variations among cells but it was argued that a quantitative difference could possibly arise if there should be a differential replication of the genes in various cells. These discussions were largely laid to rest by two types of experiments and observations beginning in the early 1950's.

BEERMAN (1952, 1962) began publishing the results of systematic developmental studies of the polytene chromosomes of the dipteran insects. The puffs at specific loci which had been viewed as the morphological basis for the presumed differences among the chromosomes in different tissues, were shown to be localized changes in bands of the polytene chromosomes associated with different developmental stages. After puffing or changing in appearance the chromosomes often returned to the original condition in a later of development. A change in the functional state of the genes could occur without a necessary change in the quantity of the DNA.

BRIGGS and KING (1952) began a series of experiments with frog eggs in which nuclei from the cells of blastula, and eventually later stages in development, were injected into enucleated eggs. The nuclei appeared to be capable of reinitiating normal development even when taken from late stages in embryos. However, the efficiency of the nuclei in reinitiating normal development decreased with age of the embryo. These studies indicated that the changes which were presumed to occur during differentiation were largely reversible. GURDON and his co-workers (GURDON 1962 a, b; GURDON and LASKEY 1970) continued these experiments with *Xenopus,* the South African frog. Most of the experiments confirmed the view that whatever changes occur during development and differentiation are reversible. While it is true that many nuclei fail to function to reinitiate development when transplanted to a frog egg, the failure could be attributed to deficiencies other than a loss or permanent change in some of the genes. However, MCKINNELL (1978), who has written a monograph on the problem, maintains that there may be non-reversible changes in the genotype during differentiation which such transplantation studies have not revealed. A more recent study of the ability of erythrocyte nuclei to support continued development when transferred to eggs of the frog, *Rana pipiens,* also leaves the question of irreversible changes of these

此为试读,需要完整PDF请访问: www.ertongbook.com