aspects of protein biosynthesis

PART A

Edited by
C. B. ANFINSEN, JR.

Aspects of Protein Biosynthesis

Edited by

C. B. Anfinsen, Jr.

LABORATORY OF CHEMICAL BIOLOGY
NIAMD
NATIONAL INSTITUTES OF HEALTH
BETHESDA, MARYLAND

Part A

ACADEMIC PRESS New York and London 1970

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

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
Berkeley Square House, London W1X 6BA

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 69-18355

PRINTED IN THE UNITED STATES OF AMERICA

List of Contributors

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

VINCENT G. ALLFREY (247), The Rockefeller University, New York, New York

CHARLES J. EPSTEIN (367), Department of Pediatrics, University of California Medical School, San Francisco, California

E. PETER GEIDUSCHEK (43), Department of Biophysics, University of Chicago, Chicago, Illinois

NICHOLAS M. KREDICH (1), Department of Medicine, Duke University Hospital, Durham, North Carolina

MARSHALL NIRENBERG (215), Laboratory of Biochemical Genetics, National Heart Institute, National Institutes of Health, Bethesda, Maryland

G. M. TOMKINS (1), Laboratory of Molecular Biology, National Institute of Arthritis and Metabolic Diseases, National Institutes of Health, Bethesda, Maryland

GUNTER VON EHRENSTEIN* (139), Department of Biophysics, Johns Hopkins University, School of Medicine, Baltimore, Maryland

* Present address: Department of Molecular Biology, Max-Planck-Institute for Experimental Medicine, Göttingen, Germany.

Preface

The term "protein biosynthesis" has come to include a very large part of what is otherwise known as "molecular biology." Indeed, in a recent, randomly selected issue of the *Proceedings of the National Academy of Sciences of the U. S.*, 17 of the 35 articles in the Biological Section dealt with some aspect of this problem. As the chapter titles of this book indicate, it is necessary to include the whole range of topics from bacterial genetics and regulatory mechanisms, through the chemistry and biology of RNA and DNA, to evolution, and the hereditary diseases of man.

The actual synthesis of peptide bonds during the elaboration of polypeptide chains remains the area of greatest uncertainty. The subjects of chain initiation, translocation and termination, as well as ribosome and polyribosome structure and formation, are undergoing intense investigation at the present time, and we may safely assume that well-informed and dependable discussions will be possible in the near future. These areas of research will form the subject matter of Part B, now in preparation, of this two-volume treatise of protein biosynthesis.

Although a definitive work on protein synthesis cannot yet be written, the group of individuals who became authors of these chapters agreed that a book that summarized those aspects now generally accepted as fact or near-fact would be of great value to students as well as to scientists working in unrelated disciplines. Speculation has been held to a minimum. A discussion of the historical development of the field has not been included since this has been treated fairly recently from a technical point of view in a volume edited by P. N. Campbell and J. R. Sargent ("Techniques in Protein Biosynthesis," Academic Press, 1967) and in a more narrative manner by R. Hendler ("Protein Biosynthesis and Membrane Biochemistry," John Wiley and Sons, 1968).

C. B. Anfinsen, Jr.

Bethesda, Maryland
November, 1969

Contents

III. Transfer RNA and Amino Acid Activation

Gunter von Ehrenstein

IV. The Flow of Information from Gene to Protein

Marshall Nirenberg

V. Biosynthetic Reactions in the Cell Nucleus

Vincent G. Allfrey

VI. The Three-Dimensional Structure and Evolution of Proteins

Charles J. Epstein

Contents of Part B (Tentative)

I

Genetic Control of Protein Structure and the Regulation of Protein Synthesis

G. M. Tomkins and Nicholas M. Kredich

I. NATURE OF THE GENETIC DETERMINANTS OF PROTEIN STRUCTURE

A. Identity of Nucleic Acids as the Genetic Substance

1. *Deoxyribonucleic Acid*

The genetic substance must have two distinct properties: a capacity for self-replication and an ability to determine cellular activity. In view of the overwhelming evidence implicating DNA in both these regards, it is difficult to realize that, until fairly recently, the chemical nature of genetic determinants was not known, and that proteins, rather than nucleic acids, seemed the most likely candidates. However, in 1944, Avery and associates (7) demonstrated conclusively that DNA, in the complete absence of protein, could carry genetic information. In these beautiful experiments, a genetic substance was isolated from one strain of pneumococcus which, when added to a culture of a second, conferred on the recipient the antigenic properties of the donor strain capsule. This transformation of the recipient persisted for many generations. An analysis of the chemical composition of the transforming substance showed it to correspond exactly to that of DNA, and its ultraviolet absorption spectrum was found to be identical to that of DNA. DNase, but not RNase or proteolytic enzymes, inactivated the transforming substance. These studies solidly established that DNA can transfer genetic

information from one cell to another. Since the recipient organism had acquired the ability to synthesize a new type of capsular polysaccharide, the authors suggested that the action of genes must be to control either enzyme formation or function.

Further proof that DNA, and not protein, is the hereditary material was presented by Hershey and Chase in a classic paper on the nature of the genetic determinants in bacteriophage T4 (*49*). This elegant work established the general mechanism by which the genes of a bacterial virus are introduced into, and replicate within, the host cell (Fig. 1). The investigators prepared phage labeled with either ^{32}P, which is incorporated into DNA, or ^{35}S, which occurs only in the viral proteins. Then phage particles, labeled in either the DNA or protein portion, were allowed to attach to bacteria for a given length of time after which the virus–host complex was separated by mechanical agitation. Under these conditions, where virus multiplication occurred normally, almost all the ^{35}S-labeled phage protein was removed from the bacterial host, but the

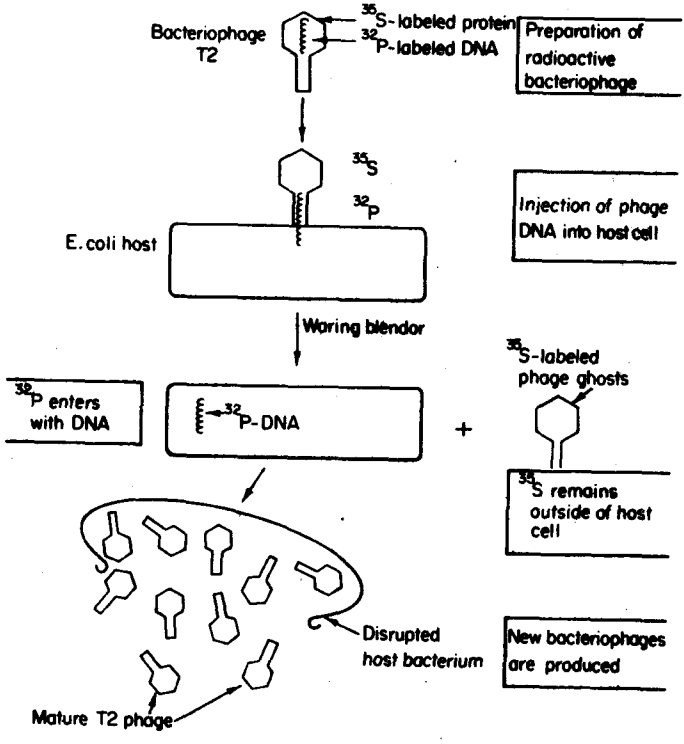

FIG. 1. Schematic representation of the Hershey–Chase experiment (*49*).

^{32}P remained. This result established that phage DNA alone can direct the synthesis of new viruses, while the proteins of the phage particle act as a kind of syringe, which injects the information-containing nucleic acid into a bacterium.

The experiments of Avery *et al.* and Hershey and Chase show that DNA carries genetic information both in bacteria and their viruses. Many more recent experiments have substantiated these conclusions and have shown that this is true in almost all living things. For example, highly purified DNA can also transform *Bacillus subtilis* (*86*). This bacterium is more suitable for genetic studies than is the *Diplococcus pneumoniae* originally used by Avery *et al.* (*7*), since it can grow on a completely defined simple medium, to permit the easy demonstration of nutritional requirements. A genetic map of the *B. subtilis* chromosome, based on transformation experiments and other techniques, has been compiled by Dubnau *et al.* (*25*). Transformation can also be shown in the classic organism of bacterial genetics *Escherichia coli*. In this case purified DNA from the lysogenic bacteriophage (lambda) can transform the cells, provided a normal phage (helper phage) is used to facilitate the entry of the nucleic acid into the bacteria (*60*). Purified DNA from certain mammalian viruses can infect animal cells with the consequent production of virus particles or other manifestations of viral infection (*24, 37*). The single-strand circular DNA molecule from the virus ϕX174 (containing 5500 nucleotides) can also be taken up and expressed in spheroplasts of *E. coli* (*46*).

Using the latter system, Goulian, Kornberg, and Sinsheimer have provided perhaps the most elegant demonstration to date that DNA carries genetic information (*42*). These workers used purified single-strand DNA [called the (+) strand], extracted from ϕX174, as a template for the *in vitro* enzymic synthesis of its complementary, (—) strand (see Chapter II), using purified *E. coli* DNA polymerase. The synthesis requires the triphosphates of the four normal deoxyribonucleosides: deoxyadenosine (dATP), deoxycytidine (dCTP), deoxyguanosine (dGTP), and deoxythymidine (dTTP). In this experiment, however, the dTTP was substituted for by its denser, synthetic analog, bromodeoxyuridine triphosphate (dBrUTP). Light circular (+) strand was used as a template, and the dense (—) complementary strand was enzymically synthesized, with its free ends joined by the recently discovered DNA ligase enzyme (*36*). The double helical complex of light template and its dense complement was dissociated, and the heavier (—) strand, synthesized *in vitro*, was isolated by centrifugation in a cesium chloride density gradient. The newly synthesized (—) strand, after separation from its (+) template, was next used as a template for the

in vitro synthesis of a light (+) strand by substituting dTTP for dBrUTP in the reaction mixture. Its free ends were joined by the ligase enzyme. In this reaction a completely synthetic duplex circular molecule was formed.

Both the separated circular (+) and (−) molecules, products of *in vitro* enzymic synthesis, as well as the synthetic duplex circle, were then used to infect spheroplasts of *E. coli*, and, in all cases, normal virus particles were formed by the host. One evident consequence of this classic experiment is that DNA indeed carries genetic information.

2. *Ribonucleic Acid*

After the general acceptance of the idea that DNA served the genetic function, it was assumed for some time that this was the only molecule that did so. The discovery of self-replicating viruses containing only RNA suggested, however, that RNA in some instances must also transmit genetic information. Fraenkel-Conrat and Williams (*29*) have clearly demonstrated the genetic role of the ribonucleic acid in the RNA-containing virus, tobacco mosaic virus (TMV). The rod-shaped virus particles consist of a single RNA chain of approximately 6400 nucleotides, surrounded by a protein coat composed of about 2100 protein units, each of molecular weight 18,000. By separating and purifying the RNA and protein components of the virus, it has been possible to demonstrate that the infectivity, and, hence, the genetic determinants of the virus, fractionates with the RNA and not the protein. The infectivity of the purified RNA is only 0.1% that of the whole virus, but can be increased several hundredfold by reconstituting the virus particle from its purified RNA and protein. The increased infectivity of the RNA–protein particle over that of the pure RNA may be attributed to the ability of the protein coat to protect the RNA from nucleases found in the tobacco leaf.

If a virus particle is reconstituted from the RNA of one strain and the coat protein of another, the progeny obtained from infection are those of the strain from which the RNA was derived (*30*). Valuable information concerning the nature of mutagenesis has been obtained by chemically modifying purified TMV-RNA, reconstituting it with normal untreated coat protein and studying the mutant progeny produced by such hybrids (*31, 93*). These reconstitution experiments strengthen the contention that nucleic acid, rather than protein, is the genetic material.

Spiegelman *et al.* (*85*) have successfully replicated the *E. coli* RNA virus Qβ in a cellfree *in vitro* system utilizing the RNA-dependent RNA replicase induced by this bacteriophage. It has been shown that this newly synthesized RNA is capable of infecting *E. coli* spheroplasts to give rise to an apparently normal bacteriophage.

Although these experiments with RNA viruses conclusively establish the ability of RNA to carry genetic determinants, DNA appears to comprise most if not all the genetic material of bacteria and higher organisms.

B. Physical and Chemical Structure of Genes and Chromosomes

1. *Structure of Deoxyribonucleic Acid*

The first account of the isolation of DNA was published in 1871 by Friedrich Miescher (*68*), who purified the substance from pus cells. By the 1940's its primary structure was generally recognized to be that of a polymer, composed of deoxyribonucleotides linked together in a linear fashion by 3′,5′-phosphodiester bonds (Fig. 2). The purines, adenine and guanine, and the pyrimidines, cytosine and thymine, appear to constitute almost the entire base content of DNA from all sources examined, except for small amounts of 5-methylcytosine in plants and animals, and 5-hydroxymethylcytosine, which occurs instead of cytosine in certain bacteriophage. The occurrence of other rare bases cannot be excluded by present analytical techniques.

In 1953, Watson and Crick (*94*) proposed a model for the secondary structure of DNA, which, now generally accepted as correct, has been a powerful stimulus to the growth of molecular genetics and our understanding of the genetic control of protein synthesis. Based on the X-ray diffraction data of Wilkins and his co-workers (*101*), and Franklin and Gosling (*32*), Watson and Crick postulated that DNA consists of two helical polynucleotide chains, each coiled about the same axis. The deoxyribose phosphate chain comprises the backbone of each helix, while the bases lie in planes perpendicular to the longitudinal axis of the molecule. While both chains are right-hand helices, the deoxyribose phosphate backbones run in opposite directions to one another with the bases from one chain sharing the same planes with bases of the opposing chain (Fig. 3). This allows specific hydrogen bonding between bases (Fig. 4) to hold the entire structure together. In the proposed structure the repeat distance between planes of nucleotide pairs is 3.4 Å, while the helix is so pitched as to repeat itself every 34 Å (every 10 nucleotide pairs); both distances were derived from X-ray diffraction data. According to the model, the diameter of the molecule is 20 Å, which also agrees with the X-ray diffraction data and with measurements performed on electron microphotographs of DNA (*9*) (Fig. 5).

From the biological standpoint, the most critical feature of this

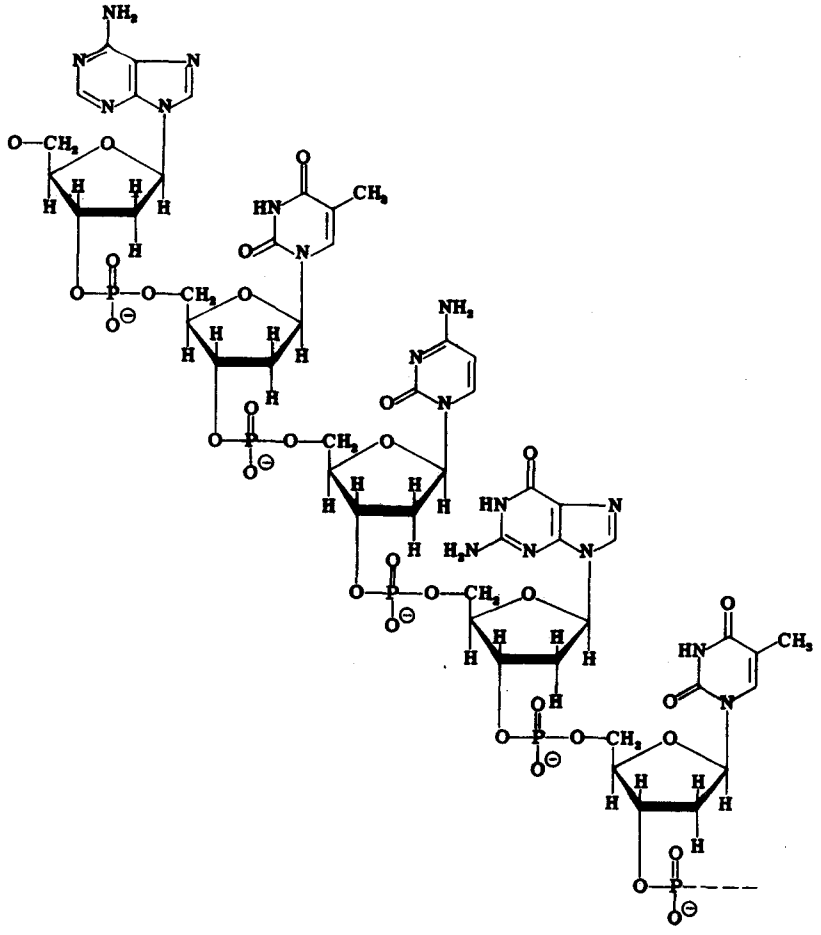

FIG. 2. The primary structure of DNA (from *60a*). Note the purine and pyrimidine bases attached to the polydeoxyribose phosphate backbone. The base sequence, adenine-thymine-cytosine-guanine-thymine, is hypothetical.

model is that adenine on one chain can base-pair only with thymine on the other chain, and guanine can pair only with cytosine (or 5-hydroxymethylcytosine or 5-methylcytosine). This is due to the fact that the distance between glycosidic bonds of opposite chains is a constant one allowing only these two possible combinations to fit within this distance. Thus, the two chains are complementary to one another in their base sequence, an A, G, G, T, C sequence in one requiring a T, C, C, A, G sequence in the other. Although the model imposes these