ESSAYS ON NUCLEIC ACIDS

ERWIN CHARGAFF

ESSAYS ON NUCLEIC ACIDS

BY

ERWIN CHARGAFF

ELSEVIER PUBLISHING COMPANY
AMSTERDAM / LONDON / NEW YORK
1963

SOLE DISTRIBUTORS FOR THE UNITED STATES AND CANADA
AMERICAN ELSEVIER PUBLISHING COMPANY, INC.
52 VANDERBILT AVENUE, NEW YORK 17, N.Y.

SOLE DISTRIBUTORS FOR GREAT BRITAIN
ELSEVIER PUBLISHING COMPANY LIMITED
12B, RIPPLESIDE COMMERCIAL ESTATE
RIPPLE ROAD, BARKING, ESSEX

LIBRARY OF CONGRESS CATALOG CARD NUMBER 63-8569

WITH 13 ILLUSTRATIONS AND 39 TABLES

Preface

Few recent advances have, for better or for worse, had such an impact on biological thinking as the discovery of base-pairing in nucleic acids. These complementariness principles do not only underlie current ideas on the structure of the nucleic acids, but they form the foundation of all speculations, more or less well-founded, on their physical properties (denaturation, hypochromicity, etc.), on the transfer of biological information from deoxyribonucleic acid to ribonucleic acid, and on the role of the latter in directing the synthesis of specific proteins. They form the basis of present explanations of the manner in which the amino acids are activated before being assembled to make a protein; they are being invoked incessantly in attempts to unravel the nucleotide code which is thought to be responsible for specifying the amino acid sequence of proteins.

It will, perhaps, surprise many readers, into whose ears a different version has been drummed for years, to learn that the first announcement of base-pairing in nucleic acids was made in an article, published early in 1950, which forms Chapter 1 of this book; the statement itself will be found on page 13. Those to whom it may sound unusually modest or restrained are asked to consider that at that time the new science of molecular biology did not yet exist.

Chapters 2 to 9 are also drawn from essays published previously at different times and at different places, some not easily accessible. They are all reprinted here without change, with the exception of Chapter 9 which underwent extensive revision. Place and year of publication are indicated in each case; and I

should like to use this occasion to thank the several publishers for the permission to reprint these articles.

The last two chapters, Nos. 10 and 11, have not been published before and are of an entirely different kind, stepping forth, as they do, without the stately periwig of a list of references. They should be considered as a *divertimento,* though not without passages in a minor key; specimens of a sort that rarely finds its way into a book dealing with scientific matters. Chapter 10 is a specimen of a recent lecture, as it was prepared for actual delivery, without undergoing the normal editing for publication which consists mostly in the substitution of references to the literature for any critical or unusual remarks that may have been made. Chapter 11 is a specimen of many conversations that I have participated in over the last few years; it is, of course, a composite of many such talks, a collage, as it were: no single person could be so dim.

There will be some, I am certain, that will find the application to scientific problems of the means of humor, of satire, and even of puns, these metaphysical hiccups of language, most unbecoming and frivolous. But there are many levels at which criticism ought to be exercised; and the critique of some of the concepts of modern science, and especially of its aberrations, has virtually disappeared at a time when it is more necessary than ever; at a time when the polarization of science has gone so far that one now "runs" for scientific awards as for a political office; that scientific lectures begin to sound like keynote speeches at political conventions; that scientific reporting has replaced the intimate gossip from Hollywood; that the persuasiveness of truth has been replaced by the strength of the acclamation; in other words, that cliques are surrounded by claques. The emergence of a Scientific Establishment, of a power elite, has given rise to a remarkable phenomenon: the appearance of what is called dogmas in biological thinking. Reason and judgment are inclined to abdicate when faced with a dogma; but they should not. Just as in political life, a stiff upper lip often conceals a soft underbelly. It is imperative that the most stringent criticism be ap-

plied to tentative scientific hypotheses that disguise themselves as dogmas. This criticism must come from within; but it can only come from an outsider at the inside.

If the title of the last chapter requires an explanation, I may quote Webster's New Collegiate Dictionary: "*amphisbaena*—a fabled serpent with a head at each end, moving either way". Whether strand separation was observed in the Middle Ages, is not recorded.

Columbia University
New York, N.Y.
December, 1962

ERWIN CHARGAFF

Contents

CHAPTER 1

*Chemical Specificity of Nucleic Acids and Mechanism of Their Enzymic Degradation**

1. INTRODUCTION

The last few years have witnessed an enormous revival in interest for the chemical and biological properties of nucleic acids, which are components essential for the life of all cells. This is not particularly surprising, as the chemistry of nucleic acids represents one of the remaining major unsolved problems in biochemistry. It is not easy to say what provided the impulse for this rather sudden rebirth. Was it the fundamental work of Hammarsten[1] on the highly polymerized deoxyribonucleic acid of calf thymus? Or did it come from the biological side, for instance, the experiments of Brachet[2] and Caspersson[3]? Or was it the very important research of Avery[4] and his collaborators on the transformation of pneumococcal types that started the avalanche?

It is, of course, completely senseless to formulate a hierarchy of cellular constituents and to single out certain compounds as more important than others. The economy of the living cell probably knows no conspicuous waste; proteins and nucleic acids, lipids and polysaccharides, all have the same importance. But one observation may be offered. It is impossible to write the history of the cell without considering its geography; and we cannot do this without attention to what may be called the chronology of the cell, *i.e.*, the sequence in which the cellular con-

* This article is based on a series of lectures given before the Chemical Societies of Zurich and Basle (June 29th and 30th, 1949), the Société de chimie biologique at Paris, and the Universities of Uppsala, Stockholm, and Milan. (Reprinted with permission from *Experientia*, 6 (1950) 201–209).

stituents are laid down and in which they develop from each other. If this is done, nucleic acids will be found pretty much at the beginning. An attempt to say more leads directly into empty speculations in which almost no field abounds more than the chemistry of the cell. Since an ounce of proof still weighs more than a pound of prediction, the important genetical functions, ascribed—probably quite rightly—to the nucleic acids by many workers, will not be discussed here. Terms such as "template" or "matrix" or "reduplication" will not be found in this lecture.

2. IDENTITY AND DIVERSITY IN HIGH-MOLECULAR CELL CONSTITUENTS

The determination of the constitution of a complicated compound, composed of many molecules of a number of organic substances, evidently requires the exact knowledge of the nature and proportion of all constituents. This is true for nucleic acids as much as for proteins or polysaccharides. It is, furthermore, clear that the value of such constitutional determinations will depend upon the development of suitable methods of hydrolysis. Otherwise, substances representing an association of many chemical individuals can be described in a qualitative fashion only; precise decisions as to structure remain impossible. When our laboratory, more than four years ago, embarked upon the study of nucleic acids, we became aware of this difficulty immediately.

The state of the nucleic acid problem at that time found its classical expression in Levene's monograph[5]. (A number of shorter reviews, indicative of the development of our conceptions concerning the chemistry of nucleic acids, should also be mentioned[6-11].) The old tetranucleotide hypothesis—it should never have been called a theory—was still dominant; and this was characteristic of the enormous sway that the organic chemistry of small molecules held over biochemistry. I should like to illustrate what I mean by one example. If in the investigation of a disaccharide consisting of two different hexoses we isolate 0.8 mole of one sugar and 0.7 mole of the other, this will be sufficient for

the recognition of the composition of the substance, provided its molecular weight is known. The deviation of the analytical results from simple, integral proportions is without importance in that case. But this will not hold for high-molecular compounds in which variations in the proportions of their several components often will provide the sole indication of the occurrence of different compounds.

In attempting to formulate the problem with some exaggeration one could say: The validity of the identification of a substance by the methods of classical organic chemistry ends with the mixed melting point. When we deal with the extremely complex compounds of cellular origin, such as nucleic acids, proteins, or polysaccharides, a chemical comparison aiming at the determination of identity or difference must be based on the nature and the proportions of their constituents, on the sequence in which these constituents are arranged in the molecule, and on the type and the position of the linkages that hold them together. The smaller the number of components of such a high-molecular compound is, the greater is the difficulty of a decision. The occurrence of a very large number of different proteins was recognized early; no one to my knowledge ever attempted to postulate a protein as a compound composed of equimolar proportions of 18 or 20 different amino acids. In addition, immunological investigations contributed very much to the recognition of the multiplicity of proteins. A decision between identity and difference becomes much more difficult when, as is the case with the nucleic acids, only few primary components are encountered. And when we finally come to high polymers, consisting of one component only, *e.g.*, glycogen or starch, the characterization of the chemical specificity of such a compound becomes a very complicated and laborious task.

While, therefore, the formulation of the tetranucleotide conception appeared explainable on historical grounds, it lacked an adequate experimental basis, especially as regards "thymonucleic acid". Although only two nucleic acids, the deoxyribose nucleic acid of calf thymus and the ribose nucleic acid of yeast, had been

examined analytically in some detail, all conclusions derived from the study of these substances were immediately extended to the entire realm of nature; a jump of a boldness that should astound a circus acrobat. This went so far that in some publications the starting material for the so-called "thymonucleic acid" was not even mentioned or that it was not thymus at all, as may sometimes be gathered from the context, but, for instance, fish sperm or spleen. The animal species that had furnished the starting material often remained unspecified.

Now the question arises: How different must complicated substances be, before we can recognize their difference? In the multiformity of its appearances nature can be primitive and it can be subtle. It is primitive in creating in a cell, such as the tubercle bacillus, a host of novel compounds, new fatty acids, alcohols, etc., that are nowhere else encountered. There, the recognition of chemical peculiarities is relatively easy. But in the case of the proteins and nucleic acids, I believe, nature has acted most subtly; and the task facing us is much more difficult. There is nothing more dangerous in the natural sciences than to look for harmony, order, regularity, before the proper level is reached. The harmony of cellular life may well appear chaotic to us. The disgust for the amorphous, the ostensibly anomalous—an interesting problem in the psychology of science—has produced many theories that shrank gradually to hypotheses and then vanished.

We must realize that minute changes in the nucleic acid, *e.g.*, the disappearance of one guanine molecule out of a hundred, could produce far-reaching changes in the geometry of the conjugated nucleoprotein; and it is not impossible that rearrangements of this type are among the causes of the occurrence of mutations.*

The molecular weight of the pentose nucleic acids, especially of those from animal tissue cells, is not yet known; and the problem of their preparation and homogeneity still is in a very sad state. But that the deoxypentose nucleic acids, prepared

---

* For additional remarks on this problem, compare Ref. 12.

under as mild conditions as possible and with the avoidance of enzymic degradation, represent fibrous structures of high molecular weight has often been demonstrated. No agreement has as yet been achieved on the order of magnitude of the molecular weight, since the interpretation of physical measurements of largely asymmetrical molecules still presents very great difficulties. But regardless of whether the deoxyribonucleic acid of calf thymus is considered as consisting of elementary units of about 35,000 which tend to associate to larger structures[13, 14] or whether it is regarded as a true macromolecule[15] of a molecular weight around 820,000, the fact remains that the deoxypentose nucleic acids are high-molecular substances which in size resemble, or even surpass, the proteins. It is quite possible that there exists a critical range of molecular weights above which two different cells will prove unable to synthesize completely identical substances. The enormous number of diverse proteins may be cited as an example. *Duo non faciunt idem* is, with respect to cellular chemistry, perhaps an improved version of the old proverb.

3. PURPOSE

We started in our work from the assumption that the nucleic acids were complicated and intricate high-polymers, comparable in this respect to the proteins, and that the determination of their structures and their structural differences would require the development of methods suitable for the precise analysis of all constituents of nucleic acids prepared from a large number of different cell types. These methods had to permit the study of minute amounts, since it was clear that much of the material would not be readily available. The procedures developed in our laboratory make it indeed possible to perform a complete constituent analysis on 2–3 mg of nucleic acid, and this in six parallel determinations.

The basis of the procedure is the partition chromatography on filter paper. When we started our experiments, only the qualitative application to amino acids was known[16]. But it was

obvious that the high and specific absorption in the ultraviolet of the purines and pyrimidines could form the basis of a quantitative ultra-micro method, if proper procedures for the hydrolysis of the nucleic acids and for the sharp separation of the hydrolysis products could be found.

4. PREPARATION OF THE ANALYTICAL MATERIAL

If preparations of deoxypentose nucleic acids are to be subjected to a structural analysis, the extent of their contamination with pentose nucleic acid must not exceed 2–3%. The reason will later be made clearer; but I should like to mention here that all deoxypentose nucleic acids of animal origin studied by us so far were invariably found to contain much more adenine than guanine. The reverse appears to be true for the animal pentose nucleic acids: in them guanine preponderates. A mixture of approximately equal parts of both nucleic acids from the same tissue, therefore, would yield analytical figures that would correspond, at least as regards the purines, to roughly equimolar proportions. Should the complete purification—sometimes an extremely difficult task—prove impossible in certain cases, one could think of subjecting preparations of both types of nucleic acid from the same tissue specimen to analysis and of correcting the respective results in this manner. This, however, is an undesirable device and was employed only in some of the preparations from liver which will be mentioned later.

It is, furthermore, essential that the isolation of the nucleic acids be conducted in such a manner as to exclude their degradation by enzymes, acid or alkali. In order to inhibit the deoxyribonucleases which require magnesium[17], the preparation of the deoxypentose nucleic acids was carried out in the presence of citrate ions[18]. It would take us here too far to describe in detail the methods employed in our laboratory for the preparation of the deoxypentose nucleic acids from animal tissues. They represent in general a combination of many procedures, as described recently for the isolation of yeast deoxyribonucleic acid[19].

In this manner, the deoxypentose nucleic acids of thymus, spleen, liver, and also yeast were prepared. The corresponding compound from tubercle bacilli was isolated *via* the nucleoprotein[20]. The procedures leading to the preparation of deoxypentose nucleic acid from human sperm will soon be published[21]. All deoxypentose nucleic acids used in the analytical studies were prepared as the sodium salts (in one case the potassium salt was used); they were free of protein, highly polymerized, and formed extremely viscous solutions in water. They were homogeneous electrophoretically and showed a high degree of monodispersity in the ultracentrifuge.

The procedure for the preparation of pentose nucleic acids from animal tissues resembled, in its first stages, the method of Clarke and Schryver[22]. The details of the isolation procedures and related experiments on yeast ribonucleic acid are as yet unpublished. Commercial preparations of yeast ribonucleic acid also were examined following purification. As has been mentioned before, the entire problem of the preparation and homogeneity of the pentose nucleic acids, and even of the occurrence of only one type of pentose nucleic acid in the cell, urgently requires re-examination.

5. SEPARATION AND ESTIMATION OF PURINES AND PYRIMIDINES

Owing to the very unpleasant solubility and polar characteristics of the purines, the discovery of suitable solvent systems and the development of methods for their quantitative separation and estimation[23, 24] presented a rather difficult problem in the solution of which Dr. Ernst Vischer had an outstanding part. The pyrimidines proved somewhat easier to handle. The choice of the solvent system for the chromatographic separation of purines and pyrimidines will, of course, vary with the particular problem. The efficiency of different solvent systems in effecting separation is illustrated schematically in Fig. 1. Two of the solvent systems listed there are suitable for the separation of the purines found in nucleic acids, *i.e.*, adenine and guanine, namely (1) *n*-butanol,

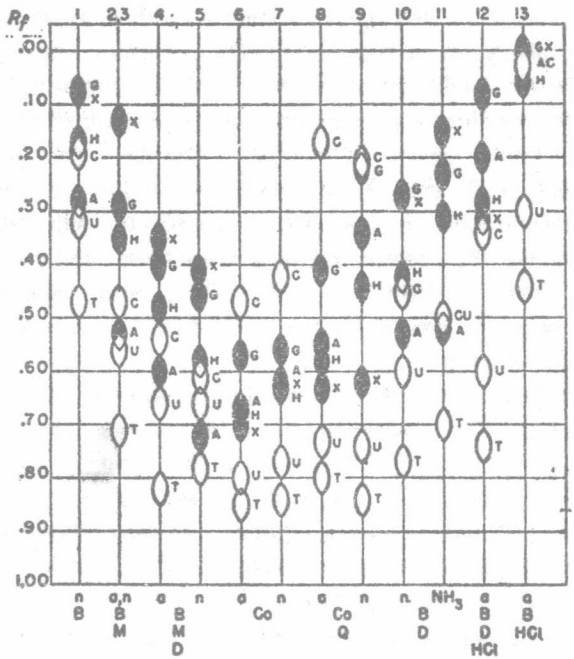

Fig. 1. Schematic representation of the position on the paper chromatogram of the purines and pyrimidines following the separation of a mixture. A = adenine, G = guanine, H = hypoxanthine, X = xanthine, U = uracil, C = cytosine, T = thymine. The conditions under which the separations were performed are indicated at the bottom: a = acidic, n = neutral, B = *n*-butanol, M = morpholine, D = diethylene glycol, Co = collidine, Q = quinoline. (Taken from E. Vischer and E. Chargaff[24].)

morpholine, diethylene glycol, water (column 5 in Fig. 1); and (2) *n*-butanol, diethylene glycol, water in an NH_3 atmosphere (column 11). The second system listed proved particularly convenient. The separation of the pyrimidines is carried out in aqueous butanol (column 1).

Following the separation, the location of the various adsorption zones on the paper must be demonstrated. Our first attempts to bring this about in ultraviolet light were unsuccessful, probably because of inadequate filtration of the light emitted by the lamp then at our disposal. For this reason, the expedient was used of fixing the separated purines or pyrimidines on the paper as

mercury complexes which then were made visible by their conversion to mercuric sulfide. The papers thus developed served as guide strips for the removal of the corresponding zones from untreated chromatograms that were then extracted and analyzed in the ultraviolet spectrophotometer. The development of the separated bases as mercury derivatives has, however, now become unnecessary, except for the preservation of permanent records, since there has for some time been available commercially an ultraviolet lamp emitting short-wave ultraviolet ("Mineralight", Ultraviolet Products Corp., Los Angeles, California). With the help of this lamp it is now easy to demonstrate directly the position of the separated purines and pyrimidines (and also of nucleosides and nucleotides[25]) which appear as dark absorption shadows on the background of the fluorescing filter paper and can be cut apart accordingly. (We are greatly indebted to Dr. C. E. Carter, Oak Ridge National Laboratory, who drew our attention to this instrument*.)

The extracts of the separated compounds are then studied in the ultraviolet spectrophotometer. The measurement of complete absorption spectra permits the determination of the purity of the solutions and at the same time the quantitative estimation of their contents. The details of the procedures employed have been published[24]. In this manner, adenine, guanine, uracil, cytosine, and thymine (and also hypoxanthine, xanthine, and 5-methylcytosine) can be determined quantitatively in amounts of 2–40 μg. The precision of the method is $\pm$ 4% for the purines and even better for the pyrimidines, if the averages of a large series of determinations are considered. In individual estimations the accuracy is about $\pm$ 6%.

Procedures very similar in principle served in our laboratory for the separation and estimation of the ribonucleosides uridine and cytidine and for the separation of deoxyribothymidine from thymine. Methods for the separation and quantitative determination of the ribonucleotides in an aqueous ammonium isobutyrate—isobutyric acid system have likewise been developed[25, 27].

* A similar arrangement was recently described[26].