MBL/WHOI

0 0301 0020669 4

ESSAYS ON NUCLEIC ACIDS

QF 55
C59

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

ALL RIGHTS RESERVED
THIS BOOK OR ANY PART THEREOF MAY NOT BE
REPRODUCED IN ANY FORM,
INCLUDING PHOTOSTATIC OR MICROFILM FORM,
WITHOUT WRITTEN PERMISSION FROM THE PUBLISHERS

TO THE MEMORY OF MY TEACHER, KARL KRAUS

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 complementarity 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 clagues. 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

<i>Preface</i>	VII
<i>Chapter 1. Chemical Specificity of Nucleic Acids and Mechanism of Their Enzymic Degradation</i>	1
1. Introduction	1
2. Identity and diversity in high-molecular cell constituents	2
3. Purpose	5
4. Preparation of the analytical material	6
5. Separation and estimation of purines and pyrimidines	7
6. Methods of hydrolysis	10
7. Composition of deoxypentose nucleic acids	11
8. Composition of pentose nucleic acids	15
9. Sugar components	18
10. Depolymerizing enzymes	18
11. Concluding remarks	20
References	23
<i>Chapter 2. Structure and Function of Nucleic Acids as Cell Constituents</i>	25
1. Biological significance of nucleic acids	26
2. Deoxypentose nucleic acids	27
3. Pentose nucleic acids	35
4. Final remarks	35
References	37
<i>Chapter 3. Deoxypentose Nucleoproteins and Their Prosthetic Groups</i>	39
1. Introduction	39
2. Deoxypentose nucleoproteins	41
3. Deoxypentose nucleic acids	50
References	60
<i>Chapter 4. The Chemistry and Function of Nucleoproteins and Nucleic Acids</i>	62
1. Introduction	62
2. Nucleoproteins	63
3. Deoxypentose nucleic acids	65
4. Fractionation of deoxypentose nucleic acids	66

5. Differences between pentose nucleic acids	69
6. Remarks on functions	70
References	75
<i>Chapter 5. The Very Big and the Very Small: Remarks on Conjugated Proteins</i>	<i>77</i>
<i>Chapter 6. Of Nucleic Acids and Nucleoproteins</i>	<i>82</i>
1. Introduction	82
2. The significance of isolated cell constituents	83
3. Two types of nucleic acid	84
4. Deoxypentose nucleic acids	86
5. Pentose nucleic acids and nucleoproteins	92
6. The meaning of regularities	93
7. Remarks on nucleoproteins	96
References	98
<i>Chapter 7. Nucleic Acids as Carriers of Biological Information</i>	<i>100</i>
1. Is there a hierarchy of cellular constituents?	100
2. On biological information chemically conveyed	102
3. Invariability of nucleic acids	104
4. Diversity of nucleic acids	105
5. Regularity of nucleic acids	106
6. Concluding remarks	107
References	108
<i>Chapter 8. First Steps towards a Chemistry of Heredity</i>	<i>109</i>
1. Introduction	109
2. Chemical basis of cellular specificity	111
3. On the alphabet of the cell and some language difficulties	113
4. Biochemical space	115
5. On the code-script of biological high polymers	117
References	125
<i>Chapter 9. The Problem of Nucleotide Sequence in Deoxy- ribonucleic Acids</i>	<i>126</i>
1. Introduction	126
2. Remarks on the conceptual basis of sequence analysis	127
3. Remarks on nomenclature	131
4. Early attempts at sequence investigation	132
5. Sequence studies through differential distribution analysis	134
6. Summarizing remarks	147
References	159
<i>Chapter 10. A Few Remarks on Nucleic Acids, Decoding, and the Rest of the World</i>	<i>161</i>
<i>Chapter 11. Amphisbaena</i>	<i>174</i>
<i>Index</i>	<i>201</i>

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 Hammarsten¹ on the highly polymerized deoxyribonucleic acid of calf thymus? Or did it come from the biological side, for instance, the experiments of Brachet² and Caspersson³? Or was it the very important research of Avery⁴ 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 monograph⁵. (A number of shorter reviews, indicative of the development of our conceptions concerning the chemistry of nucleic acids, should also be mentioned⁶⁻¹¹.) 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 macromolecule¹⁵ 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 known¹⁶. 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 magnesium¹⁷, the preparation of the deoxypentose nucleic acids was carried out in the presence of citrate ions¹⁸. 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 acid¹⁹.

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 nucleoprotein²⁰. The procedures leading to the preparation of deoxypentose nucleic acid from human sperm will soon be published²¹. 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 Schryver²². 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. Chargaff²⁴.)

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 nucleotides²⁵) 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 published²⁴. 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 μ g. 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 described²⁶.

6. METHODS OF HYDROLYSIS

It has long been known that the purines can be split off completely by a relatively mild acid hydrolysis of the nucleic acids. This could be confirmed in our laboratory in a more rigorous manner by the demonstration that heating at 100° for 1 hour in *N* sulfuric acid effects the quantitative liberation of adenine and guanine from adenylic and guanylic acids, respectively. The liberation of the pyrimidines, however, requires much more energetic methods of cleavage. Heating at high temperatures with strong mineral acid under pressure is usually resorted to. To what extent these procedures brought about the destruction of the pyrimidines could not be ascertained previously owing to the lack of suitable analytical procedures. The experiments summarized in Table 1, which are quoted from a recent paper²³, show that the extremely robust cleavage methods with mineral

TABLE 1

RESISTANCE OF PYRIMIDINES TO TREATMENT WITH STRONG ACID*

Experiment No.	Acid	Heat- ing time	Concentration shift, % of starting concentration		
		(min)	Uracil	Cytosine	Thymine
1	HCl (10%)	90	+ 62	— 63	+ 3
2	10 <i>N</i> HCOOH +	60	+ 3	— 5	0
3	<i>N</i> HCl (1 : 1)	120	+ 24	— 19	0
4	HCOOH (98–100%)	60	0	— 1	— 2
5		120	0	+ 2	+ 1

* A mixture of pyrimidines of known concentration was dissolved in the acids indicated below and heated at 175° in a bomb tube. The concentration shifts of the individual pyrimidines were determined through a comparison of the recoveries of separated pyrimidines before and after the heating of the mixture.

acids usually employed must have led to a very considerable degradation of cytosine to uracil. Uracil and also thymine are

much more resistant. For this reason, we turned to the hydrolysis of the pyrimidine nucleotides by means of concentrated formic acid. For the liberation of the purines, *N* sulfuric acid (100°, 1 h) is employed; for the liberation of the pyrimidines, the purines are first precipitated as the hydrochlorides by treatment with dry HCl gas in methanol and the remaining pyrimidine nucleotides cleaved under pressure with concentrated formic acid (175°, 2 h). This procedure proved particularly suitable for the investigation of the deoxypentose nucleic acids. For the study of the composition of pentose nucleic acids a different procedure, making use of the separation of the ribonucleotides, was developed more recently, which will be mentioned later.

7. COMPOSITION OF DEOXYPENTOSE NUCLEIC ACIDS

It should be stated at the beginning of this discussion that the studies conducted thus far have yielded no indication of the occurrence in the nucleic acids examined in our laboratory of unusual nitrogenous constituents. In all deoxypentose nucleic acids investigated the purines were adenine and guanine, the pyrimidines cytosine and thymine. The occurrence in minute amounts of other bases, *e.g.*, 5-methylcytosine, can, however, not yet be excluded. In the pentose nucleic acids uracil occurred instead of thymine.

A survey of the composition of deoxyribose nucleic acid extracted from several organs of the ox is provided in Table 2 (Ref. 29). The molar proportions reported in each case represent averages of several hydrolysis experiments. The composition of deoxypentose nucleic acids from human tissues is similarly illustrated in Table 3 (Ref. 30). The preparations from human liver were obtained from a pathological specimen in which it was possible, thanks to the kind cooperation of Dr. M. Faber, to separate portions of unaffected hepatic tissue from carcinomatous tissue consisting of metastases from the sigmoid colon, previous to the isolation of the nucleic acids.

In order to show examples far removed from mammalian

TABLE 2
COMPOSITION OF DEOXYRIBONUCLEIC ACID OF OX
(in moles of nitrogenous constituent per mole of P)

<i>Constituent</i>	<i>Thymus</i>			<i>Spleen</i>		<i>Liver</i>
	<i>Prep. 1</i>	<i>Prep. 2</i>	<i>Prep. 3</i>	<i>Prep. 1</i>	<i>Prep. 2</i>	
Adenine	0.26	0.28	0.30	0.25	0.26	0.26
Guanine	0.21	0.24	0.22	0.20	0.21	0.20
Cytosine	0.16	0.18	0.17	0.15	0.17	
Thymine	0.25	0.24	0.25	0.24	0.24	
Recovery	0.88	0.94	0.94	0.84	0.88	

TABLE 3
COMPOSITION OF DEOXPENTOSE NUCLEIC ACID OF MAN
(in moles of nitrogenous constituent per mole of P)

<i>Constituent</i>	<i>Sperm</i>		<i>Thymus</i>	<i>Liver</i>	
	<i>Prep. 1</i>	<i>Prep. 2</i>		<i>Normal</i>	<i>Carcinoma</i>
Adenine	0.29	0.27	0.28	0.27	0.27
Guanine	0.18	0.17	0.19	0.19	0.18
Cytosine	0.18	0.18	0.16		0.15
Thymine	0.31	0.30	0.28		0.27
Recovery	0.96	0.92	0.91		0.87

organs, the composition of two deoxyribonucleic acids of microbial origin, namely from yeast¹⁹ and from avian tubercle bacilli²⁰, is summarized in Table 4 (Ref. 31).

The very far-reaching differences in the composition of deoxypentose nucleic acids of different species are best illustrated by a comparison of the ratios of adenine to guanine and of thymine to cytosine as given in Table 5. It will be seen that in all cases where enough material for statistical analysis was available highly significant differences were found. The analytical figures on which Table 5 is based were derived by comparing the ratios found for individual nucleic acid hydrolysates of one species regardless of the organ from which the preparation

TABLE 4
COMPOSITION OF TWO MICROBIAL DEOXYRIBONUCLEIC ACIDS

<i>Constituent</i>	<i>Yeast</i>		<i>Avian tubercle bacilli</i>
	<i>Prep. 1</i>	<i>Prep. 2</i>	
Adenine	0.24	0.30	0.12
Guanine	0.14	0.18	0.28
Cytosine	0.13	0.15	0.26
Thymine	0.25	0.29	0.11
Recovery	0.76	0.92	0.77

was isolated. This procedure assumes that there is no organ specificity with respect to the composition of deoxypentose nucleic acids of the same species. That this appears indeed to be the case may be gathered from Tables 2 and 3 and even better from Table 6 where the average purine and pyrimidine ratios in individual tissues of the same species are compared. That the isolation of nucleic acids did not entail an appreciable fractionation is shown by the finding that when whole defatted human spermatozoa, after being washed with cold 10% trichloroacetic acid, were analyzed, the same ratios of adenine to guanine and of thymine to cytosine were found as are reported in Tables 5 and 6. It should also be mentioned that all preparations, with the exception of those from human liver, were derived from pooled starting material representing a number, and in the case of human spermatozoa a very large number, of individuals.

The deoxypentose nucleic acids extracted from different species thus appear to be different substances or mixtures of closely related substances of a composition constant for different organs of the same species and characteristic of the species.

The results serve to disprove the tetranucleotide hypothesis. It is, however, noteworthy—whether this is more than accidental, cannot yet be said—that in all deoxypentose nucleic acids examined thus far the molar ratios of total purines to total pyrimidines, and also of adenine to thymine and of guanine to cytosine, were not far from 1.

TABLE 5
MOLAR PROPORTIONS OF PURINES AND PYRIMIDINES IN DEOXYPENTOSE NUCLEIC ACIDS FROM DIFFERENT SPECIES

Species	Number of different organs	Number of different preparations	Adenine/Guanine			Thymine/Cytosine		
			Number of hydrolyses ^a	Mean ratio	Standard error	Number of hydrolyses ^a	Mean ratio	Standard error
Ox ^b	3	7	20	1.29	0.013	6	1.43	0.03
Man ^c	2	3	6	1.56	0.008	5	1.75	0.03
Yeast	1	2	3	1.72	0.02	2	1.9	
Avian tubercle bacillus	1	1	2	0.4		1	0.4	

^a In each hydrolysis between 12 and 24 determinations of individual purines and pyrimidines were performed.

^b Preparations from thymus, spleen, and liver served for the purine determinations, the first two organs for the estimation of pyrimidines.

^c Preparations from spermatozoa and thymus were analysed.

TABLE 6

MOLAR PROPORTIONS OF PURINES AND PYRIMIDINES IN DEOXYPENTOSE
NUCLEIC ACIDS FROM DIFFERENT ORGANS OF ONE SPECIES

<i>Species</i>	<i>Organ</i>	<i>A/G</i>	<i>T/C</i>
Ox	Thymus	1.3	1.4
	Spleen	1.2	1.5
	Liver	1.3	
Man	Thymus	1.5	1.8
	Sperm	1.6	1.7
	Liver (normal)	1.5	1.8
	Liver (carcinoma)	1.5	1.8

Abbreviations: A = adenine, G = guanine, T = thymine, C = cytosine.

8. COMPOSITION OF PENTOSE NUCLEIC ACIDS

Here a sharp distinction must be drawn between the prototype of all pentose nucleic acid investigations—the ribonucleic acid of yeast—and the pentose nucleic acids of animal cells. Nothing is known as yet about bacterial pentose nucleic acids. In view of the incompleteness of our information on the homogeneity of pentose nucleic acids, which I have stressed before, I feel that the analytical results on these preparations do not command the same degree of confidence as do those obtained for the deoxy-pentose nucleic acids.

Three procedures, to which reference is made in Tables 7 (Refs. 25, 28) and 8, were employed in our laboratory for the analysis of pentose nucleic acids.

In *Procedure 1*, the pentose nucleic acid was hydrolyzed to the nucleotide stage with alkali, at pH 13.5 and 30°, and the nucleotides, following adjustment to about pH 5, separated by chromatography with aqueous ammonium isobutyrate–isobutyric acid as the solvent. Under these conditions, guanylic acid shares its position on the chromatogram with uridylic acid; but it is possible to determine the concentrations of the two components in the eluates by simultaneous equations based on the ultraviolet

TABLE 7
COMPOSITION OF YEAST RIBONUCLEIC ACIDS
(in moles of nitrogenous constituent per mole of P)

<i>Constituent</i>	<i>Preparation 1</i>			<i>Preparation 2</i>			<i>Preparation 3</i>		
	<i>Proc.</i> 1	<i>Proc.</i> 2	<i>Proc.</i> 3	<i>Proc.</i> 1	<i>Proc.</i> 2	<i>Proc.</i> 3	<i>Proc.</i> 1	<i>Proc.</i> 2	<i>Proc.</i> 3
Adenylic acid	0.29	0.26	0.26	0.27		0.24	0.25	0.23	0.24
Guanylic acid	0.28	0.29	0.26	0.25		0.25	0.26	0.28	0.26
Cytidylic acid	0.18	0.17	0.24	0.20			0.21	0.21	
Uridylic acid	0.20	0.20	0.08	0.18	0.19		0.20	0.25	
Recovery	0.95	0.92	0.84	0.90			0.92	0.97	

absorption of the pure nucleotides^{25, 27}. The very good recoveries of nucleotides obtained in terms of both nucleic acid phosphorus and nitrogen show the cleavage by mild alkali treatment of pentose nucleic acids to be practically quantitative.

In *Procedure 2*, the purines are first liberated by gaseous HCl in dry methanol and the evaporation residue of the reaction mixture is adjusted to pH 13.5 and then treated as in *Procedure 1*. In this manner, uridylic and cytidylic acids, adenine and guanine are separated and determined on one chromatogram.

The determinations of free purines and pyrimidines in acid hydrolysates of pentose nucleic acids, following the methods outlined before for the deoxypentose nucleic acids, are listed as *Procedure 3*. It will be seen that it is mainly uracil which in this procedure escapes quantitative determination. This is due to the extreme refractoriness of uridylic acid to complete hydrolysis by acids, a large portion remaining partially unsplit as the nucleoside uridine. As matters stand now, I consider the values for purines yielded by *Procedures 1* and *3* and those for pyrimidines found by *Procedures 1* and *2* as quite reliable.

A survey of the composition of yeast ribonucleic acid is provided in Table 7. Preparations 1 and 2, listed in this table, were commercial preparations that had been purified in our laboratory and had been subjected to dialysis; Preparation 3 was isolated

from baker's yeast by Magasanik in this laboratory by procedures similar to those used for the preparation of pentose nucleic acids from animal tissues and had not been dialyzed. It will be seen that the results are quite constant and not very far from the proportions required by the presence of equimolar quantities of all four nitrogenous constituents.

An entirely different picture, however, was encountered when the composition of pentose nucleic acids from animal cells was investigated. A preliminary summary of the results, in all cases obtained by Procedure 1, is given in Table 8. Here guanylic acid was the preponderating nucleotide followed, in this order, by cytidylic and adenylic acids; uridylic acid definitely was a minor constituent. This was true not only of the ribonucleic acid of pancreas which has been known to be rich in guanine³²⁻³⁴, but also of all pentose nucleic acids isolated by us from the livers of three different species (Table 8).

TABLE 8
COMPOSITION OF PENTOSE NUCLEIC ACIDS FROM ANIMAL TISSUES

<i>Constituent</i>	<i>Calf liver</i>	<i>Ox liver</i>	<i>Sheep liver</i>	<i>Pig liver</i>	<i>Pig pancreas</i>
Guanylic acid	16.3	14.7	16.7	16.2	22.5
Adenylic acid	10	19	10	10	10
Cytidylic acid	11.1	10.9	13.4	16.1	9.8
Uridylic acid	5.3	6.6	5.6	7.7	4.6
Purines:pyrimidines	1.6	1.4	1.4	1.1	2.3

In the absence of a truly reliable standard method for the isolation of pentose nucleic acid from animal tissue, generalizations are not yet permitted; but it would appear that pentose nucleic acids from the same organ of different species are more similar to each other, at least in certain respects (*e.g.*, the ratio of guanine to adenine), than are those from different organs of the same species. (Compare the pentose nucleic acids from the liver and the pancreas of pig in Table 8.)

9. SUGAR COMPONENTS

It is deplorable that such designations as deoxyribose and ribose nucleic acids continue to be used as if they were generic terms. Even the "thymus nucleic acid of fish sperm" is encountered in the literature. As a matter of fact, only in a few cases have the sugars been identified, namely, *d*-2-deoxyribose as a constituent of the guanine and thymine nucleosides of the deoxypentose nucleic acid from calf thymus, D-ribose as a constituent of the pentose nucleic acids from yeast, pancreas, and sheep liver.

Since the quantities of novel nucleic acids usually will be insufficient for the direct isolation of their sugar components, we attempted to employ the very sensitive procedure of the filter paper chromatography of sugars³⁵⁻³⁷ for the study of the sugars isolated from minute quantities of nucleic acids. It goes without saying that identifications based on behavior in adsorption or partition are by no means as convincing as the actual isolation, but they will at least permit a tentative classification of new nucleic acids. Thus far the pentose nucleic acids of pig pancreas²⁸ and of the avian tubercle bacillus³¹ have been shown to contain ribose, the deoxypentose nucleic acids of ox spleen²⁹, yeast and avian tubercle bacilli³¹ deoxyribose. It would seem that the free play with respect to the variability of components that nature permits itself is extremely restricted where nucleic acids are concerned.

10. DEPOLYMERIZING ENZYMES

Enzymes capable of bringing about the depolymerization of both types of nucleic acids have long been known; but it is only during the last decade that crystalline ribonuclease³⁸ and deoxyribonuclease³⁹ from pancreas have become available thanks to the work of Kunitz. Important work on the latter enzyme was also done by McCarty¹⁸.

We were, of course, interested in applying the chromatographic micromethods for the determination of nucleic acid

TABLE 9

ENZYMIC DEGRADATION OF CALF THYMUS DEOXYRIBONUCLEIC ACID

<i>Diges- tion (hours)</i>	<i>Dialysis (hours)</i>	<i>Distribution of fractions (% of original)</i>	<i>Composition of fractions (molar proportions)</i>				
			<i>A/G</i>	<i>T/C</i>	<i>A/C</i>	<i>Pu/Py</i>	
Original	0	0	100	1.2	1.3	1.6	1.2
Dialysate	6	6	53	1.2	1.2	1.2	1.0
Dialysis residue	24	72	7	1.6	2.2	3.8	2.0

Abbreviations: A = adenine, G = guanine, T = thymine, C = cytosine, Pu = purines, Py = pyrimidines.

constituents to studies of enzymic reaction mechanisms for which they are particularly suited. The action of crystalline deoxyribonuclease on calf thymus deoxyribonucleic acid resulted in the production of a large proportion of dialyzable fragments (53% of the total after 6 hours digestion), without liberation of ammonia or inorganic phosphate. But even after extended digestion there remained a non-dialyzable core whose composition showed a significant divergence from both the original nucleic acid and the bulk of the dialysate. The preliminary findings summarized in Table 9 (Ref. 40) indicate a considerable increase in the molar proportions of adenine to guanine and especially to cytosine, of thymine to cytosine, and of purines to pyrimidines. This shows that the dissymmetry in the distribution of constituents, found in the original nucleic acid (Table 2), is intensified in the core. The most plausible explanations of this interesting phenomenon, the study of which is being continued, are that the preparations consisted of more than one deoxypentose nucleic acid or that the nucleic acid contained in its chain clusters of nucleotides (relatively richer in adenine and thymine) that were distinguished from the bulk of the molecule by greater resistance to enzymic disintegration.

In this connection another study, carried out in collaboration with Zamenhof, should be mentioned briefly that dealt with the

deoxypentose nuclease of yeast cells^{41, 42}. This investigation afforded a possibility of exploring the mechanisms by which an enzyme concerned with the disintegration of deoxypentose nucleic acid is controlled in the cell. Our starting point again was the question of the specificity of deoxypentose nucleic acids; but the results were entirely unexpected. Since we had available a number of nucleic acids from different sources, we wanted to study a pair of deoxypentose nucleic acids as distant from each other as possible, namely that of the ox and that of yeast, and to investigate the action on them of the two deoxypentose nucleases from the same cellular sources. The deoxyribonuclease of ox pancreas has been thoroughly investigated, as was mentioned before. Nothing was known, however, regarding the existence of a yeast deoxypentose nuclease.

It was found that fresh salt extracts of crushed cells contained such an enzyme in a largely inhibited state, due to the presence of a specific inhibitor protein. This inhibitor specifically inhibited the deoxypentose nuclease from yeast, but not that from other sources, such as pancreas. The yeast enzyme depolymerized the deoxyribose nucleic acids of yeast and of calf thymus, which differ chemically, as I have emphasized before, at about the same rate. In other words, the enzyme apparently exhibited inhibitor specificity, but not substrate specificity. It is very inviting to assume that such relations between specific inhibitor and enzyme, in some ways reminiscent of immunological reactions, are of more general biological significance. In any event, a better understanding of such systems will permit an insight into the delicate mechanisms through which the cell manages the economy of its life, through which it maintains its own continuity and protects itself against agents striving to transform it.

11. CONCLUDING REMARKS

Generalizations in science are both necessary and hazardous; they carry a semblance of finality which conceals their essentially provisional character; they drive forward, as they retard; they

add, but they also take away. Keeping in mind all these reservations, we arrive at the following conclusions. The deoxypentose nucleic acids from animal and microbial cells contain varying proportions of the same four nitrogenous constituents, namely, adenine, guanine, cytosine, thymine. Their composition appears to be characteristic of the species, but not of the tissue, from which they are derived. The presumption, therefore, is that there exists an enormous number of structurally different nucleic acids; a number, certainly much larger than the analytical methods available to us at present can reveal.

It cannot yet be decided, whether what we call the deoxypentose nucleic acid of a given species is one chemical individual, representative of the species as a whole, or whether it consists of a mixture of closely related substances, in which case the constancy of its composition merely is a statistical expression of the unchanged state of the cell. The latter may be the case if, as appears probable, the highly polymerized deoxypentose nucleic acids form an essential part of the hereditary processes; but it will be understood from what I said at the beginning that a decision as to the identity of natural high polymers often still is beyond the means at our disposal. This will be particularly true of substances that differ from each other only in the sequence, not in the proportion, of their constituents. The number of possible nucleic acids having the same analytical composition is truly enormous. For example, the number of combinations exhibiting the same molar proportions of individual purines and pyrimidines as the deoxyribonucleic acid of the ox is more than 10^{56} , if the nucleic acid is assumed to consist of only 100 nucleotides; if it consists of 2,500 nucleotides, which probably is much nearer the truth, then the number of possible "isomers" is not far from 10^{1500} .

Moreover, deoxypentose nucleic acids from different species differ in their chemical composition, as I have shown before; and I think there will be no objection to the statement that, as far as chemical possibilities go, they could very well serve as one of the agents, or possibly as the agent, concerned with the transmis-

sion of inherited properties. It would be gratifying if one could say—but this is for the moment no more than an unfounded speculation—that just as the deoxypentose nucleic acids of the nucleus are species-specific and concerned with the maintenance of the species, the pentose nucleic acids of the cytoplasm are organ-specific and involved in the important task of differentiation.

I should not want to close without thanking my colleagues who have taken part in the work discussed here; they are, in alphabetical order, Miss R. Doniger, Mrs. C. Green, Dr. B. Magasanik, Dr. E. Vischer, and Dr. S. Zamenhof.

REFERENCES

- ¹ E. HAMMARSTEN, *Biochem. Z.*, 144 (1924) 383.
- ² J. BRACHET, in *Nucleic Acid, Symposia Soc. Exptl. Biol.*, 1 (1947) 207; cf. J. BRACHET, in *Nucleic Acids and Nucleoproteins, Cold Spring Harbor Symposia Quant. Biol.*, 12 (1947) 18.
- ³ T. CASPERSSON, in *Nucleic Acid, Symposia Soc. Exptl. Biol.*, 1 (1947) 127.
- ⁴ O. T. AVERY, C. M. MACLEOD AND M. MCCARTY, *J. Exptl. Med.*, 79 (1944) 137.
- ⁵ P. A. LEVENE AND L. W. BASS, *Nucleic Acids*, Chemical Catalog Co., New York, 1931.
- ⁶ H. BREDERECK, *Fortschr. Chem. org. Naturstoffe*, 1 (1938) 121.
- ⁷ F. G. FISCHER, *Naturwissenschaften*, 30 (1942) 377.
- ⁸ R. S. TIPSON, *Advances in Carbohydrate Chem.*, 1 (1945) 193.
- ⁹ J. M. GULLAND, G. R. BARKER AND D. O. JORDAN, *Ann. Rev. Biochem.*, 14 (1945) 175.
- ¹⁰ E. CHARGAFF AND E. VISCHER, *Ann. Rev. Biochem.*, 17 (1948) 201.
- ¹¹ F. SCHLENK, *Advances in Enzymol.*, 9 (1949) 455.
- ¹² E. CHARGAFF, in *Nucleic Acids and Nucleoproteins, Cold Spring Harbor Symposia Quant. Biol.*, 12 (1947) 28.
- ¹³ E. HAMMARSTEN, *Acta Med. Scand.*, Suppl. 196 (1947) 634.
- ¹⁴ G. JUNGNER, I. JUNGNER AND L.-G. ALLGÉN, *Nature*, 163 (1949) 849.
- ¹⁵ R. CECIL AND A. G. OGSTON, *J. Chem. Soc.*, (1948) 1382.
- ¹⁶ R. CONSDEN, A. H. GORDON AND A. J. P. MARTIN, *Biochem. J.*, 38 (1944) 224.
- ¹⁷ F. G. FISCHER, I. BÖTTGER AND H. LEHMANN-ECHTERNACHT, *Z. physiol. Chem.*, *Hoppe-Seyler's*, 271 (1941) 246.
- ¹⁸ M. MCCARTY, *J. Gen. Physiol.*, 29 (1946) 123.
- ¹⁹ E. CHARGAFF AND S. ZAMENHOF, *J. Biol. Chem.*, 173 (1948) 327.
- ²⁰ E. CHARGAFF AND H. F. SAIDEL, *J. Biol. Chem.*, 177 (1949) 417.
- ²¹ S. ZAMENHOF, L. B. SHETTLES AND E. CHARGAFF, *Nature*, 165 (1950) 756.
- ²² G. CLARKE AND S. B. SCHRYVER, *Biochem. J.*, 11 (1917) 319.
- ²³ E. VISCHER AND E. CHARGAFF, *J. Biol. Chem.*, 168 (1947) 781.
- ²⁴ E. VISCHER AND E. CHARGAFF, *J. Biol. Chem.*, 176 (1948) 703.
- ²⁵ E. CHARGAFF, B. MAGASANIK, R. DONIGER AND E. VISCHER, *J. Am. Chem. Soc.*, 71 (1949) 1513.
- ²⁶ E. R. HOLIDAY AND E. A. JOHNSON, *Nature*, 163 (1949) 216.
- ²⁷ E. VISCHER, B. MAGASANIK AND E. CHARGAFF, *Federation Proc.*, 8 (1949) 263.
- ²⁸ E. VISCHER AND E. CHARGAFF, *J. Biol. Chem.*, 176 (1948) 715.
- ²⁹ E. CHARGAFF, E. VISCHER, R. DONIGER, C. GREEN AND F. MISANI, *J. Biol. Chem.*, 177 (1949) 405.
- ³⁰ E. CHARGAFF, S. ZAMENHOF AND C. GREEN, *Nature*, 165 (1950) 756.
- ³¹ E. VISCHER, S. ZAMENHOF AND E. CHARGAFF, *J. Biol. Chem.*, 177 (1949) 429.

- ³² E. HAMMARSTEN, *Z. physiol. Chem., Hoppe-Seyler's*, 109 (1920) 141.
- ³³ P. A. LEVENE AND E. JORPES, *J. Biol. Chem.*, 86 (1930) 389.
- ³⁴ E. JORPES, *Biochem. J.*, 28 (1934) 2102.
- ³⁵ S. M. PARTRIDGE, *Nature*, 158 (1946) 270.
- ³⁶ S. M. PARTRIDGE AND R. G. WESTALL, *Biochem. J.*, 42 (1948) 238.
- ³⁷ E. CHARGAFF, C. LEVINE AND C. GREEN, *J. Biol. Chem.*, 175 (1948) 67.
- ³⁸ M. KUNITZ, *J. Gen. Physiol.*, 24 (1940) 15.
- ³⁹ M. KUNITZ, *Science*, 108 (1948) 19.
- ⁴⁰ S. ZAMENHOF AND E. CHARGAFF, *J. Biol. Chem.*, 178 (1949) 531.
- ⁴¹ S. ZAMENHOF AND E. CHARGAFF, *Science*, 108 (1948) 628.
- ⁴² S. ZAMENHOF AND E. CHARGAFF, *J. Biol. Chem.*, 180 (1949) 727.

*Structure and Function of Nucleic Acids as Cell Constituents**

It is safe to say that living systems require the presence of both types of nucleic acid or, in the case of parasitic systems, the presence of at least one. If, to use a designation of Schrödinger¹, one refers to the chromosomes as "the hereditary code-script", the great biological importance of all components of these nuclear structures, *viz.*, nucleic acids, proteins, and, perhaps, lipids, is obvious, unless one assumes that one or the other of these components has been added by nature as a meaningless and purely decorative flourish. This is, however, not likely.

In animals and higher plants the deoxypentose nucleic acids (DNA) are exclusively or almost exclusively situated in the nucleus. Pentose nucleic acids (PNA) are present in the nucleoli and the various cytoplasmic elements, *e.g.*, the mitochondria, submicroscopic particles, etc. Recent work with Elson² on the nucleotide composition of PNA in different fractions of rat liver cells has provided preliminary evidence of differences in composition between nuclear and cytoplasmic PNA. It is not yet known whether this will be generally true; but results, recently reported by Magasanik³, seem to point to the presence in pig liver PNA of two differently composed fractions.

Whether DNA really is limited to the nucleus is not entirely certain, since the available cytochemical or cytophysical methods presumably require the presence of a compact mass of DNA and may not reveal its occurrence in a diffusely distributed form, as could be the case in the cytoplasm of egg cells (*cf.* Ref. 4).

* Reprinted with permission from *Federation Proc.*, 10 (1951) 654-659.

The bacterial cell represents a special case. Whether the microbial nucleus is the sole repository of DNA in microorganisms cannot yet be decided with certainty.

1. BIOLOGICAL SIGNIFICANCE OF NUCLEIC ACIDS

The discussion of the biological significance of a ubiquitous cell constituent is, strictly speaking, superfluous. But there exist a few important instances pointing to a direct involvement of nucleic acids; and some of them will be listed here briefly.

All virus preparations so far described are, or contain, nucleoproteins (*cf.* the recent survey of Davidson⁵). The same seems to be true of intracellular parasites, such as rickettsiae^{6, 7} or paramecin⁸.

Specific DNA preparations are known which are able to induce the transformation of bacterial types. This extremely important phenomenon, first discovered in pneumococci⁹, has later been shown to operate also in *E. coli*¹⁰ and in *Hemophilus influenzae*¹¹ (*cf.* also Ref. 12). The possibility that reactions of this kind are of more general biological importance and not limited to the field of bacterial transformations cannot be rejected. What appears particularly remarkable is that it is here the free nucleic acid and not a nucleoprotein (as in the case of viruses) that is able to impose its own synthesis on the receptor cell, whereas in general nucleic acids seem to occur in cells only in the form of conjugated nucleoproteins. The mechanisms through which these transformations take place and the chemical features distinguishing these biologically active DNA specimens are completely obscure. There is, however, little doubt that it is the bacterial DNA itself, or a particular DNA fraction present in the transforming preparations, which is the carrier of activity. Recent work on the agent operative in the transformation of *H. influenzae* has shown that highly purified DNA preparations from two types are active in extremely low concentrations: 0.0004 μg of DNA per ml in type b; 0.01 μg in type c¹³.

The investigation of the relative efficiencies of different wave-

lengths of ultraviolet light has yielded curves closely resembling, but not entirely identical with, the UV absorption spectrum of nucleic acids. This work, mainly due to Stadler and Uber, Hollaender and Emmons, and Knapp and Schreiber, has been reviewed by Lea¹⁴.

A certain degree of constancy, within the same species, of the DNA concentration per diploid nucleus, and of roughly one-half this amount per haploid sperm nucleus, has been discovered by Boivin and his collaborators^{4, 15, 16} and confirmed in other laboratories^{17, 18}.

DNA is in its composition identical in different tissues of the same species¹⁹. Moreover, in the few cases where comparison was possible, no chemical differences have been observed between the composition of the DNA from the sperm cells and from differentiated tissues of the same species, in contrast to the very different composition of nuclear proteins in such instances.

2. DEOXYPENTOSE NUCLEIC ACIDS

All DNA preparations that have been studied in detail have several features in common. They are asymmetrical molecules of high molecular weight (around 10^6), yielding extremely viscous solutions. They appear to contain the same deoxy sugar, namely 2-deoxyribose. They contain two purines, adenine and guanine, and two or three pyrimidines, *viz.* thymine and cytosine, and in several cases²⁰ also 5-methylcytosine. The deoxyribonucleotides released enzymically from calf thymus DNA appear to be the 5'-phosphates^{21, 22}.

The conclusions to which our work has led us have been summarized recently^{19, 23}, and I shall limit myself here only to the main points.

a. DNA is in its composition characteristic of the species from which it is derived. This can in many, but not in all, cases be demonstrated by determining the ratios in which the individual purines and pyrimidines occur. There will, however, be very many borderline cases in which such differences in composition

are not sufficiently significant to permit their use as the sole criterion of differentiation. The most important question of the sequence in which, in a particular nucleic acid, the nucleotides follow each other has so far barely been approached. The elaboration of methods for sequence analysis is, perhaps, one of the most urgent problems in nucleic acid chemistry, since differences in nucleotide sequence may very well be among the determinants of chemical and biological specificity.

b. No differences in composition have so far been found in DNA from different tissues of the same species. This provisional conclusion refers only to the over-all composition. It is in this connection noteworthy that no chemical differences appear to exist between the composition of DNA from normal human tissue²⁴ and that of preparations from human cancer tissue²⁵.

c. The tetranucleotide hypothesis is incorrect.

d. There exist a number of regularities. Whether these are merely accidental cannot yet be decided. In almost all DNA preparations studied until now the ratio of total purines to total pyrimidines never was far from 1. Similarly the ratios of adenine to thymine and of guanine to cytosine were near 1.

e. There appear to exist two main groups of DNA, namely the "AT type", in which adenine and thymine predominate, and the "GC type", in which guanine and cytosine are the major constituents. The latter has so far been found only in certain microorganisms²⁶.

Data to support these conclusions have, in the main, been presented in previous publications and were summarized recently^{19, 23}. I should like to limit myself here to the discussion of a few as yet unpublished results. The study of the composition of the DNA of salmon sperm²⁷ has brought out some of the regularities mentioned before particularly clearly (Tables 10 and 11). This substance belongs to the "AT type".

Another investigation, undertaken in collaboration with G. Brawerman, deals with the DNA of wheat germ and the course of its degradation by crystalline pancreatic deoxyribonuclease. Previous work in our laboratory on the enzymic disintegration

TABLE 10

SALMON SPERM DNA; PROPORTIONS
(in moles of nitrogenous constituent per mole of P in hydrolysate)

<i>Constituent</i>	<i>Mean proportion</i>	<i>Standard error</i>
Adenine	0.280	0.005
Guanine	0.196	0.004
Cytosine	0.192	0.006
Thymine	0.274	0.005

TABLE 11

SALMON SPERM DNA; MOLAR RELATIONSHIPS

Adenine to guanine	1.43
Thymine to cytosine	1.43
Adenine to thymine	1.02
Guanine to cytosine	1.02
Purines to pyrimidines	1.02
P accounted for as percentage of P in hydrolysate	95.8 ($\pm$ 1.6)
Average number of gram-atoms N per mole constituent	3.7
Atomic N : P ratio in DNA preparations	3.6, 3.7

TABLE 12

WHEAT GERM DNA; INTACT PREPARATION AND ENZYMICALLY PRODUCED CORES
(in moles of nitrogenous constituent per mole of P in hydrolysate)

<i>Constituent</i>	<i>Intact DNA</i>	<i>19% core</i>	<i>8% core</i>
Adenine	0.27	0.33	0.35
Guanine	0.22	0.20	0.20
Cytosine	0.16	0.12	0.10
5-Methylcytosine	0.06	0.04	0.04
Thymine	0.27	0.26	0.23
Total purines	0.49	0.53	0.55
Total pyrimidines	0.49	0.42	0.37
Recovery	0.98	0.95	0.92

TABLE 13

MOLAR RATIOS IN DNA PREPARATIONS OF DIFFERENT ORIGIN

Source	Adenine to guanine	Thymine to cytosine	Adenine to thymine	Guanine to cytosine	Purines to pyrim- idines	Amino groups to enolic hydroxyls
Ox	1.29	1.43	1.04	1.00	1.1	1.4
Man	1.56	1.75	1.00	1.00	1.0	1.3
Hen	1.45	1.29	1.06	0.91	0.99	1.5
Salmon	1.43	1.43	1.02	1.02	1.02	1.4
Wheat	1.22	1.18*	1.00	0.97*	0.99	1.4
Yeast	1.67	1.92	1.03	1.20	1.0	1.3
<i>Hemophilus influenzae</i> , type c	1.74	1.54	1.07	0.91	1.0	1.5
<i>E. coli</i> K-12	1.05	0.95	1.09	0.99	1.0	1.6
Avian tubercle bacillus	0.4	0.4	1.09	1.08	1.1	1.7
<i>Serratia marcescens</i>	0.7	0.7	0.95	0.86	0.9	1.6
Hydrogen organism Bacillus Schatz	0.7	0.6	1.12	0.89	1.0	1.7

* In these computations the sum of cytosine and methylcytosine was used. If cytosine alone is considered, the thymine to cytosine ratio is 1.62 and that of guanine to cytosine 1.33.

of calf thymus DNA²⁸ had shown it to proceed according to a peculiar and complex pattern. The action of the enzyme resulted in the formation of dialyzable fragments and of a dialysis residue (core). The latter was characterized by greatly increased ratios of adenine to guanine, thymine to cytosine, purines to pyrimidines, and by greater resistance to enzymic attack. Preliminary results of studies of this type on wheat germ DNA are presented in Table 12. The analytical findings on the intact DNA are in good agreement with analyses recently reported for the same nucleic acid from two other laboratories^{29, 30}. The figures given here for the "19% core" and the "8% core" refer to the dialysis residues recovered when 81 and 92% of the DNA respectively had been converted by the enzyme to dialyzable fragments. The trend of degradation appears similar to that observed with calf thymus DNA.

The molar ratios found in the DNA specimens studied in our laboratory are compared in Table 13. The tendency toward certain regularities will be observed. The figures for hen DNA (chicken erythrocytes) and for the DNA from the K-12 strain of *E. coli* must be considered as preliminary. Both components were studied in collaboration with B. Gandelman. The DNA of *Hemophilus influenzae*, type c, was studied by S. Zamenhof. A few points are noteworthy. In the case of wheat germ DNA, methylcytosine and cytosine apparently must be considered together, if the regular ratios, observed in most other instances, are to be obtained. Another remarkable fact is that the regularities are maintained even in the nucleic acids of the "GC type" despite the complete inversion in individual ratios.

As was already mentioned before, it is almost impossible to decide at present whether these regularities are entirely fortuitous or whether they reflect the existence in all DNA preparations of certain common structural principles, irrespective of far-reaching differences in their individual composition and the absence of an easily recognizable periodicity. It may be assumed that the nucleic acids as we know them today, as for that matter also the proteins, are the result of an age-long selection process in the

course of which many less suitable or less stable components must have been eliminated. One could speak of the survival of the fittest nucleic acids. Such macromolecules will exhibit diversity and uniformity at the same time, since they are called upon to perform, in diverse species, the same tasks. It is, therefore, perhaps not astonishing that the nucleic acids will share certain features that may be directly connected with their stability or their ability to form conjugated nucleoproteins. One property which is quite striking is the uniform absorption spectrum in the ultraviolet of all highly polymerized DNA specimens, both with respect to the position and the intensity of the absorption maximum. The center of absorption fluctuates only between 257 and 261 $m\mu$ and the (ϵ) P is around 6600. Another surprising feature is the balance between amino groups and enolic hydroxyls in all DNA preparations examined by us. (See last column of Table 13.) Even in the core preparations, this ratio changed only very little.

But let us return for a moment to the other outstanding characteristic of nucleic acids, *viz.*, their diversity. If we accept the evidence of the existence of species-specific DNA, then there arise many new questions, both of a biological and chemical nature. DNA presumably is an important part of the chromosomes and may be surmised to be involved in their biological functions. Does this mean that a cell contains as many different DNA individuals as it contains genes? Or can one and the same species-specific DNA form so many three-dimensional structures, in connection with the proteins to which it is attached, that the genic requirements are fulfilled? (For a more detailed discussion of some of these points, *cf.* Ref. 31.) This question, as so many others in this field, cannot yet be answered. No way has as yet been found to fractionate a family of very similar macromolecules which may differ in no more than the sequence of a few of their component nucleotides. But one could perhaps say that the more regular the arrangement of nucleotides is in a given DNA, the less the chance of its forming many different specific structures. And this brings us again to the very important question of

nucleotide sequence and periodicity in a nucleic acid chain. Any simplified assumption with respect to periodicity has been disproved by the studies on the course of action of deoxyribonuclease on the DNA of calf thymus²⁸ and of wheat germ. The composition of both the dialyzable degradation products and the dialysis residues, the "cores", exhibited continuous and characteristic changes with respect to the distribution of purines and pyrimidines. One must conclude that the sequence is highly aperiodic and that it is not inconceivable that the same cellular DNA could give rise to many different nucleoproteins, depending upon the shape and configuration of the particular protein.

It must, moreover, be understood that the recognition of periodicity, *i.e.*, the presence of recurring units, will be particularly difficult in a macromolecule of the type of DNA. If, for instance, in a chain composed of 3000 nucleotides a particular sequence of 100 consecutive nucleotides were repeated 30 times, this periodicity could not be recognized, unless we had a method producing cleavage only at the points where these repeating units are joined. In other words, the perception of periodicity would require the proper distance for a bird's-eye view which will not be easy to attain.

Another approach to the problem of sequence analysis in DNA may be seen in the study of its controlled chemical degradation. That the purines can be detached from a nucleic acid with much greater ease than can the pyrimidines has long been known; but the resulting end product, thymic acid, was rather nondescript. In collaboration with M. E. Hodes and C. Tamm it has been possible to develop procedures, soon to be published, in which all purines could be cleaved from a DNA preparation, leaving behind a non-dialyzable product in a yield of about 94% that retained all the pyrimidines of the original DNA in unchanged proportions. We have designated preparations of this type as apurinic acid. These compounds may prove of interest for structural studies on DNA, since the position in the DNA chain of the initially present purine nucleotides is now marked by reactive aldehydo groups. The properties of a typical preparation of

TABLE 14
APURINIC ACID FROM CALF THYMUS DNA

	DNA	Apurinic acid
N	14.9	6.0
P	9.1	10.7
N/P	3.6	1.2
Absorption maximum, $m\mu^*$	258.5	268
Absorption maximum, ϵ (P)	6,600	4,600
Optical rotation**	+ 102°	+ 50°
Adenine, moles per mole P	0.28	0
Guanine, moles per mole P	0.21	0
Cytosine, moles per mole P	0.21	0.18
Thymine, moles per mole P	0.26	0.24
Thymine to cytosine	1.3	1.3

* In *M* phosphate buffer, pH 7.1.

** In 0.1 *M* phosphate buffer, pH 7.1.

TABLE 15
ACTION OF RIBONUCLEASE ON PNA

Source	Fraction	Nucleotide composition (as moles/100 <i>M</i> nucleotide in starting material)				Total
		Guanylic acid	Adenylic acid	Cytidylic acid	Uridylic acid	
Yeast	Starting	30	21	23	26	100
	Dialyzable (total nucleotides)	18	15	19	26	78
	Dialyzable (mononucleotides)	0	0	14	19	33
	Core	66	18	8	8	
Pig liver	Starting	36	19	29	16	100
	Dialyzable (total nucleotides)	18	14	26	15	73
	Dialyzable (mononucleotides)	0	0	18	10	28
	Core	71	15	0	14	

apurinic acid are contrasted in Table 14 with those of the calf thymus DNA specimen that served as the starting material.

3. PENTOSE NUCLEIC ACIDS

Time does not permit an adequate discussion of the chemistry of PNA. In the past few years, significant contributions to this field were made by Gulland, Kerr, Schmidt, Allen, Loring, Cohn, and many other workers. I should like to make brief mention of some recent work done by Magasanik in our laboratory in which the course of action of crystalline ribonuclease on PNA from yeast and from pig liver was studied with the use of chromatographic and spectroscopic procedures. About 60–70%, or in some cases somewhat more, of the nucleotides present in the initial substrates were liberated by enzymic action as rapidly dialyzable nucleotides of low molecular weight. This fraction consisted of free cytidylic and uridylic acids, which comprised a high proportion of the total pyrimidine nucleotides, and of combined purine and pyrimidine nucleotides. All PNA preparations yielded a non-dialyzable residue, resistant to enzymic attack, which was found to consist to about two thirds of guanylic acid and of varying amounts of the other nucleotides. A selection of these experiments is presented in Table 15. These and other findings have led us to certain conceptions concerning the specificity of ribonuclease and the structure of PNA; but reference must be made to a more detailed article³.

4. FINAL REMARKS

It is fitting to conclude this all too sketchy survey with a confession of ignorance. What our studies have taught us more than anything else is how little we know as yet about the chemistry of nucleic acids. The chemical specificity of macromolecules and the interactions between them through which the organization of the cell is maintained can only partly be understood in terms of our present knowledge. In the approach to a scientific problem

two principles are operative: generalization and simplification. Both are necessary and both dangerous. It is obvious that we can learn more geometry from the illustrations in a textbook on projective geometry than from the beautiful pictures in Sir D'Arcy Thompson's *On Growth and Form*. But it is difficult to say where the danger line lies beyond which oversimplification will produce a dogmatic ignorance. Should we stress the multiformity of nature, which makes us forget the simplicity of its basic designs; or should the essential shape win over the accidental form? In Wycherley's *The Country Wife* a quack is addressed as follows: "Doctor, thou wilt never make a good chemist, thou art so incredulous and impatient". If patience and credulity were all the chemist needed, the problem of the nucleic acids—still so baffling and elusive—would have been solved a long time ago.

REFERENCES

- ¹ E. SCHRÖDINGER, *What is Life?*, Cambridge University Press, London, 1945, p. 19.
- ² D. ELSON AND E. CHARGAFF, *Federation Proc.*, 10 (1951) 180.
- ³ B. MAGASANIK AND E. CHARGAFF, *Biochim. Biophys. Acta*, 7 (1951) 396.
- ⁴ C. VENDRELY AND R. VENDRELY, *Compt. rend. Soc. Biol.*, 143 (1949) 1386.
- ⁵ J. N. DAVIDSON, *The Biochemistry of the Nucleic Acids*, Methuen, London, 1950, p. 136.
- ⁶ S. S. COHEN AND E. CHARGAFF, *J. Biol. Chem.*, 154 (1944) 691.
- ⁷ S. S. COHEN, *J. Immunol.*, 65 (1950) 475.
- ⁸ W. J. VAN WAGTENDONK, *J. Biol. Chem.*, 173 (1948) 691.
- ⁹ O. T. AVERY, C. M. MACLEOD AND M. MCCARTY, *J. Exptl. Med.*, 79 (1944) 137.
- ¹⁰ A. BOIVIN, A. DELAUNAY, R. VENDRELY AND Y. LEHOULT, *Experientia*, 1 (1945) 334; 2 (1946) 139.
- ¹¹ H. E. ALEXANDER AND G. LEIDY, *J. Exptl. Med.*, 93 (1951) 345.
- ¹² M. U. DIANZANI, *Boll. ist. sieroterap. milan.*, 29 (1950) 161.
- ¹³ S. ZAMENHOF, G. LEIDY, H. E. ALEXANDER, P. L. FITZGERALD AND E. CHARGAFF, *Arch. Biochem. Biophys.*, 40 (1952) 50.
- ¹⁴ D. E. LEA, *Actions of Radiations on Living Cells*, Cambridge University Press, London, 1947, p. 181.
- ¹⁵ A. BOIVIN, R. VENDRELY AND C. VENDRELY, *Compt. rend.*, 226 (1948) 1061.
- ¹⁶ R. VENDRELY AND C. VENDRELY, *Experientia*, 4 (1948) 434; 5 (1949) 327.
- ¹⁷ A. E. MIRSKY AND H. RIS, *Nature*, 163 (1949) 666.
- ¹⁸ C. LEUCHTENBERGER, R. VENDRELY AND C. VENDRELY, *Proc. Natl. Acad. Sci. U.S.*, 37 (1951) 33.
- ¹⁹ E. CHARGAFF, *Experientia*, 6 (1950) 201. [See Chapter 1 of this book.]
- ²⁰ G. R. WYATT, *Nature*, 166 (1950) 237.
- ²¹ E. VOLKIN, J. X. KHYM AND W. E. COHN, *J. Am. Chem. Soc.*, 73 (1951) 1533.
- ²² C. E. CARTER, *J. Am. Chem. Soc.*, 73 (1951) 1537.
- ²³ E. CHARGAFF, *J. Cellular Comp. Physiol.*, 38, Suppl. 1 (1951) 41.
- ²⁴ E. CHARGAFF, S. ZAMENHOF AND C. GREEN, *Nature*, 165 (1950) 756.
- ²⁵ E. CHARGAFF, *Abstr. 5th Intern. Cancer Congr., Paris*, 1950, p. 101.
- ²⁶ E. CHARGAFF, S. ZAMENHOF, G. BRAWERMAN AND L. KERIN, *J. Am. Chem. Soc.*, 72 (1950) 3825.
- ²⁷ E. CHARGAFF, R. LIPSHITZ, C. GREEN AND M. E. HODES, *J. Biol. Chem.*, 192 (1951) 223.
- ²⁸ S. ZAMENHOF AND E. CHARGAFF, *J. Biol. Chem.*, 178 (1948) 531; 187 (1950) 1.

- ²⁹ M. M. DALY, V. G. ALLFREY AND A. E. MIRSKY, *J. Gen. Physiol.*, 33 (1950) 497.
- ³⁰ G. R. WYATT, *Biochem. J.*, 47 (1950) vii.
- ³¹ E. CHARGAFF, *Cold Spring Harbor Symposia Quant. Biol.*, 12 (1947) 28.

CHAPTER 3

*Deoxypentose Nucleoproteins and Their Prosthetic Groups**

1. INTRODUCTION

Some time in 1800, the German poet Kleist wrote a letter to a friend in which he mentioned an idea that had occurred to him when passing through an arched gateway: "Why, I thought, does the vault not collapse, though entirely without support? It stands, I replied, because all the stones want to fall down at the same time." Similarly, one could say that the cell lives because all its components desire to die simultaneously. They survive, in other words, by getting into each other's way.

That branch of chemistry that is, among many other things, entrusted with conducting a traffic survey of the living cell is comparatively young. At the present time, it is true of cytochemistry, as essentially of many other sciences, that we can recognize the very general and the very special; but what lies between—and that is the greater part of the world—is completely obscure.

In trying to compare the state of biochemistry, for example, of fifty years ago with that of today, one could use a metaphor. In the old times, the knowledge of biology was perhaps similar to what could be made out in a very large, very dark house. Many objects could be more felt than seen with equal dimness, once the eyes got used to the darkness; and scientists were conscious of the limiting conditions under which they worked. In our time, however, a few very powerful and very narrow beams of light have been thrown into a few corners of this dark house,

* Reprinted with permission from *Symposia Soc. Exptl. Biol.*, 9 (1955) 32-47.

and several things can be seen in a clarity and illumination that almost distort their significance. But at the same time we have lost our dark-adaptation; and since we all have a tendency to follow the light, we have moved into these cosy corners, to the detriment of the rest, which still is, by far, the major part of nature. In pointing this out one runs the risk of being accused of trying to spread darkness.

As concerns the subject of the present conference, it will become clear that the biological significance of fibrous proteins can best be defined in those instances to which I have referred above as very special, for instance, in muscular contraction or in blood coagulation. With regard to other conjugated proteins, among which must be counted the nucleoproteins, it is their very ubiquity that has made difficult the recognition of their function, unless one is satisfied with the trite assignment of an essentially mechanical task; but this amounts to no more than saying that these proteins are, because they are. What could, of course, be questioned is whether a purely teleological search for "functions" is likely to prove adequate for the future development of our science. While substances and reactions endure, interpretations often perish with unseemly rapidity; and few things have as short a half-life time as the premature assignment of a function.

One more introductory remark: When speaking of nucleoproteins as fibrous proteins, one must specify that this designation refers principally to the deoxypentose nucleoproteins, though some fibrous pentose nucleoproteins are known. Furthermore, there is little indication that the carrier proteins themselves are fibrous. This character is conferred on the conjugated protein by the shape of the prosthetic group it contains, namely, the deoxypentose nucleic acids. And, strictly speaking, why these nucleic acids, or for that matter many other macromolecules of natural occurrence, exist as fibers of so marked an anisometry cannot yet be explained satisfactorily on chemical grounds.

In the following pages a brief survey of the chemistry of those nucleoproteins that can be classified as fibrous, namely, the deoxypentose nucleoproteins, will be followed by the consider-

ation of their prosthetic groups, the deoxypentose nucleic acids. It will not be possible to do more than mention some of the current problems. For a fuller account, compare a previous survey¹. Reviews of the chemistry of the pentose nucleoproteins have recently been published^{2, 3}.

2. DEOXYPENTOSE NUCLEOPROTEINS

a. Classification

The deoxynucleoproteins are usually defined as conjugated proteins in which the deoxypentose nucleic acid, functioning as a prosthetic group, is linked to the protein by electrostatic attraction or by secondary valence forces. It is, however, extremely difficult to distinguish, in macromolecules, between these two types of combination. It is, in addition, equally difficult to decide whether the nucleoprotein isolated in a given case really pre-existed in the cell or whether it was the result of a fortuitous combination of solutes present in the same extract. A genuine nucleoprotein, which in some of its properties ought to differ from a mere mixture of its components, must be regarded as a geometrically unique combination of two giant polyampholytes. It is only recently that attempts have been made to define more sharply the distinction between a nucleoprotein and a protein nucleate artifact⁴.

Since the deoxypentose nucleic acid fraction of a given nucleus is composed of a large number of differently constituted individuals⁵⁻⁷, a nucleoprotamine or a nucleohistone must comprise many different conjugated proteins. A search of the literature will lead to the conclusion that the number of genuine deoxypentose nucleoproteins isolated until now is small. It can, moreover, not yet be stated whether the major part, or all, of the deoxypentose nucleic acids present in the cell occurs in the form of nucleoproteins.

As regards the protein moiety, three principal groups of deoxynucleoproteins may be distinguished: (1) the nucleoprotamines occurring in ripe spermatozoa of many fish genera; (2) the

nucleohistones mostly investigated in the form of calf thymus nucleohistone; (3) the nucleoproteins in the proper sense of this term in which the nucleic acid is combined with a protein lacking the basic properties of the protamines and histones⁸. We are, however, very far from a satisfactory classification of the types of deoxynucleoprotein occurring in cell nuclei, especially with regard to changes in the composition of the protein moiety taking place during development and in the various phases of the mitotic cycle.

b. Isolation

The principal methods for the preparation of nucleoprotamines and nucleohistones, *i.e.*, complexes in which the protein partner carries a positive charge, are based on (a) extraction with strong salt solutions⁹; (b) extraction with solutions of low ionic strength¹⁰. The first-mentioned procedure, widely used as the first step in the isolation of deoxypentose nucleic acids, is accompanied by an, at least partial, dissociation and may lead to artifacts. It has, on the other hand, the advantage that the activity of deoxyribonucleases is suppressed in strong salt solutions. By extraction with solutions of low ionic strength the exposure of the conjugated protein to dissociating conditions in the course of its preparation is avoided. There exists, however, the danger of a partial enzymic degradation of the nucleic acid brought about by the release of nucleases. As examples of the isolation of nucleohistone at a low electrolyte concentration the studies on calf thymus nucleohistone in Refs. 4 and 11 may be cited; the isolation of a nucleoprotamine from sea-urchin sperm is discussed in Ref. 12. Preparations afforded by extraction with strong salt solutions are described, for instance, in Ref. 13. The isolation from avian tubercle bacilli of a nucleoprotein with entirely different solubility properties has been reported⁸.

c. Properties

A real decision on the native state or the attributes of intactness of a nucleoprotein will hardly be possible before suitable

biological test procedures have been found. The nucleoproteins of calf thymus and of fish sperm remain the only readily accessible deoxynucleoproteins, and most of our knowledge is derived from the study of these compounds. Nucleohistone is soluble at a very low ionic strength. As the electrolyte concentration increases, the solubility first drops rapidly, reaching a minimum in isotonic solutions, and then rises gradually. Though variations are found, freshly prepared nucleohistone is, in general, soluble in NaCl solutions stronger than 0.7 or 0.8 *M*. Small amounts of phosphorylated impurities can be extracted from nucleohistone by 0.02–0.15 *M* NaCl solutions; but they usually lack the typical nucleic acid spectrum (see Table 16). The solubility properties of nucleoprotamine seem to be essentially similar.

The absorption spectrum in the ultraviolet of a nucleoprotein is, in general, identical with that of its nucleic acid moiety (maximum around 260 $m\mu$). No depression of the extinction coefficient of the nucleic acid component is noticeable, nor does the extent of extinction change materially under conditions of almost complete dissociation (compare Table 16). This is of interest, since the extinction of a nucleoprotein could have been expected to be less than that of the free nucleic acid, if the purines and pyrimidines of the latter were involved in the architecture of the intact conjugated protein in such a manner as to bring about the suppression of some of the chromophores¹⁴.

The nucleic acid content of different nucleoproteins varies from about 35 to 60% of the dry weight. In most nucleohistone preparations it is around 50%. In trout and herring nucleoprotamines a phosphorus to arginine ratio of 1 has been found¹⁵.

Although there exists considerable evidence that deoxypentose nucleoproteins, or at any rate the nucleoprotamines and nucleohistones, occur in strong salt solutions in a largely dissociated state, in which a far-reaching separation between nucleic acid and protein has taken place, it would probably be wrong to regard such solutions as representing no more than a mixture of these components.

TABLE 16
PROPERTIES OF CALF THYMUS NUCLEOHISTONE SOLUTIONS AT DIFFERENT NaCl CONCENTRATIONS

Prep. No.	Properties of supernatant fluids*	Molarity of NaCl in supernatant fluids							
		0	0.02	0.05	0.50	0.65	0.80	1.0	4.3
1	<i>P</i>	18.7	2.2	2.0	3.7	14.7	18.3	17.9	19.6
	Viscosity	1,100	—	—	—	300	400	550	950
	Extinction	6,250	5,450	5,550	5,600	5,500	6,650	6,500	6,650
2	<i>P</i>	16.9	2.1	2.0	3.3	12.5	17.2	16.7	17.7
	Viscosity	1,600	—	—	—	250	400	600	950
	Extinction	6,450	6,600	6,550	6,050	6,200	6,650	6,600	6,700

* The figures for *P* are in $\mu\text{g/ml}$. The values for viscosity refer to η_{sp} , (*P*), i.e., the specific viscosity divided by the molarity of the solution with respect to *P*; those for extinction to $\epsilon(P)_{260}$, i.e., the atomic extinction coefficient with respect to *P*, determined at 260 $\text{m}\mu$. (Taken from Crampton *et al.*⁴.)

d. Dissociation and cleavage

Before the discussion of the nucleic acids proper a few words should be said about the cleavage and dissociation of the nucleoproteins, especially the nucleohistones and nucleoprotamines. The cleaving effect of high electrolyte concentrations was first observed by Bang¹⁶; it is of great importance for the isolation of reasonably intact deoxypentose nucleic acids.

A systematic study in our laboratory of the behavior of nucleoproteins towards electrolytes has opened a promising approach to the problem of the fractionation of deoxypentose nucleic acids. When calf thymus nucleohistone is exposed to increasing NaCl concentrations under conditions permitting the removal of both undissociated nucleoprotein and liberated protein, increasing quantities of sodium nucleate will be found in solution^{4, 17}. Similar results are obtained when the time of contact with the salt solution is varied rather than the concentration of the latter¹⁸. When the removal of the non-dissociated portion is brought about in the presence of a denaturing phase,

Fig. 2. Fractional dissociation of calf thymus nucleohistone (curve A) and of wheat germ nucleoprotein (curve B) in the presence of a CHCl_3 phase. The quantities of nucleic acid P (as percentage of P in the starting material) extracted at the NaCl concentrations indicated on the abscissa are plotted as the ordinate.

Fig. 3. Fractional dissociation, in the presence of a CHCl_3 phase, of artificial complexes between the deoxyribonucleic acids of calf thymus (curves A and B) and pig liver (curve C) with globin (curve A) or histone (curves B and C). Co-ordinates as in Fig. 2. (Taken from Crampton *et al.*⁶.)

such as chloroform-pentanol, a large number of distinct nucleic acid fractions can be separated, as will be discussed below. The fractional dissociation of calf thymus nucleohistone and of the nucleoprotein of wheat germ, based on previous work⁴ and on unpublished experiments with Miss Rakoma Lipshitz, is shown in Fig. 2. These are curves taken from single experiments; but the typical shape can be reproduced easily.

e. Behavior of artifacts

The behavior of artificially prepared complexes of nucleic acids with positively charged substances has also been studied. Such nucleate compounds have been prepared not only with histone and globin⁶, but also with synthetic polyelectrolytes, *e.g.*, polylysine or polyvinylamine¹⁹. The fractional dissociation of deoxyribonucleate complexes with globin or histone is illustrated in

Fig. 3. In general, globin nucleates appear to undergo dissociation more readily than histone nucleates. The relative proportions of histone and nucleic acid and the state of polymerization of the latter are of great importance for the properties of the artifacts (Fig. 4). It can be seen that the ease of dissociation of the complexes increases with the relative amount of nucleic acid present and decreases very markedly with the viscosity of the nucleic acid. It may be mentioned that under appropriate conditions nucleohistone is rendered soluble at a low ionic strength by the addition of large amounts of degraded deoxyribonucleic acid²⁰.

Fig. 4. Successive extraction with NaCl solutions, in the presence of a CHCl_3 phase, of artificial histone complexes with calf thymus deoxyribonucleic acid. In (a) the amount of histone varied, in (b) the degree of polymerization of the nucleic acid. The weight ratio of histone to nucleic acid P was 23 (curve 1), 18 (curve 2), 14 (curve 3), 4.5 (curve 4), 17 (curves 5-7). The viscosities, as $\eta_{sp}/(P)$, i.e., the specific viscosity divided by the molarity of the solution with respect to P, were: curve 5, 1000 (intact preparation); curve 6, 420 (0.1% solution heated at 75° for 5 min); curve 7, 240 (0.1% solution heated at 75° for 40 min). Co-ordinates as in Fig. 2. (Taken from Crampton *et al.*⁶)

The cationic species used also influences the progress of the dissociation. The remarkable effect of an admixture of as little as 0.06 M MgCl_2 to the NaCl solutions is shown in Fig. 5. When

solutions of MgCl_2 alone are employed, the liberation of nucleic acid from its complex with histone occurs at even lower salt concentrations, as much as 70% of the nucleic acid P being detached by 0.3 M MgCl_2 and almost the entire amount by a 0.6 M solution.

Fig. 5. Effect of magnesium ions on the fractional dissociation, in the presence of a CHCl_3 phase, of artificial histone complexes with calf thymus deoxyribonucleic acid. The weight ratio of histone to nucleic acid P was 17. The successive extractions were performed with NaCl solutions of the indicated molarity in the absence (curve 1) and in the presence of 0.06 M MgCl_2 (curve 2). Co-ordinates as in Fig. 2. (Taken from Cramp-ton *et al.*⁶.)

In these experiments the histone nucleate complex was formed at the lowest NaCl concentration employed for fractionation. The conditions under which the complex is formed are, however, not without influence on its subsequent behavior. This is brought out by the experiments assembled in Fig. 6. Two differently treated nucleohistone preparations are compared in curves 1 and 2. Before being precipitated by dialysis against 0.15 M NaCl, one portion was in contact with 3 M NaCl (curve 2), the other portion remained in distilled water (curve 1). The precipitates then were extracted successively with salt solutions of increasing strength in the absence of a denaturing chloroform phase. Similarly, artificial histone nucleates were formed in the absence of salt

Fig. 6. Successive extraction with NaCl solutions of increasing strength, in the absence (curves 1-4) and in the presence (curves 5 and 6) of a CHCl_3 phase, of calf thymus nucleohistone (curves 1 and 2) or of artificial histone complexes (curves 3-6) with calf thymus deoxyribonucleic acid. Compare the text for experimental details. Co-ordinates as in Fig. 2.

(Taken from unpublished experiments with Miss Rakoma Lipshitz.)

(curve 3) and in 3 M NaCl (curve 4) and extracted successively. In another series, similarly produced complexes (curve 5 without, curve 6 with 3 M NaCl) were extracted stepwise in the presence of CHCl_3 -pentanol. These experiments permit a comparison of the behavior of nucleohistone and of histone nucleate complexes, prepared at a low electrolyte concentration (curves 1 and 3), towards extraction with salt solutions in the absence of chloroform with that of preparations that had initially been in contact with a 3 M NaCl solution (curves 2 and 4). The effect of chloroform on the rate of appearance of nucleic acid P in the extracts is demonstrated by a comparison of curve 3 with 6 and of curve 4 with 5. Certain characteristic differences between nucleohistone and histone nucleate are brought out when curve 1 is compared with 3 and curve 2 with 4.

f. Isolation of deoxypentose nucleic acids

The first step in the preparation of a deoxypentose nucleic acid is, irrespective of whether this is done consciously or not, the isolation of a nucleoprotein. Once the nucleoprotein or a mixture of nucleic acid and protein has been extracted, the removal of the protein is carried out by treatment with chloroform, saturation with sodium chloride, or with the aid of an anionic detergent. The several methods have been reviewed critically in a recent survey¹.

3. DEOXYPENTOSE NUCLEIC ACIDS

One of the surest signs that one has reached a stage of truth is that its affirmation begins to bore the audience. I shall, therefore, limit myself to a brief consideration of the principal conclusions regarding the composition and structure of nucleic acids, many of which already had been formulated six or seven years ago²¹, *i.e.*, in what were the early days of modern nucleic acid research.

a. Is there more than one deoxypentose nucleic acid in nature?

All evidence points to the existence of a very large number of differently constituted nucleic acids; a number probably exceeding by far what can be revealed by constituent analysis. For it should be remembered that, while differences in nucleotide composition signify divergences in nucleotide sequence, an identical composition does not prove an identical sequence. An early estimate of the possible number of isomers having the composition of calf thymus deoxyribonucleic acid, which assumed a chain of 2,500 nucleotides, gave the figure of 10^{1500} . Although the differentiation between nucleic acids on the basis of differences in the contents of their nitrogenous constituents is a relatively crude expedient, since it does not permit a distinction between sequence variations that are not accompanied by a change in total composition, it can be shown in many cases that nucleic acids of different origin differ in composition (compare

the survey in Ref. 1). To mention only the most extreme examples: the ratio of the sum of adenine and thymine to that of guanine and cytosine is about 1.9 in the deoxyribonucleic acid of *Echinocardium cordatum*, but about 0.4 in that of avian tubercle bacilli. It is, in fact, possible to make a gross distinction between three types of deoxypentose nucleic acid: (a) the AT type, in which adenine and thymine are the major constituents; (b) an intermediate type, in which all bases are present in nearly equimolar proportions (found only in *Escherichia coli*); and (c) the GC type, in which guanine and cytosine predominate. While, therefore, the question whether there exists more than one deoxypentose nucleic acid must be answered in the affirmative, a decision whether different tissues of the same host yield entirely identical nucleic acid specimens is more difficult. For the moment, it would appear that differences in the composition of preparations from different organs, if they exist at all, are of very doubtful significance.

b. Do there exist recognizable repeating units?

The tetranucleotide hypothesis is incorrect. One must conclude that there is no subunit of recognizably recurrent structure larger than a mononucleotide. If repeating polynucleotide units occur in the nucleic acid chain, we have no way as yet of demonstrating their existence. The nucleic acids must be regarded as complicated and intricate high polymers of undetermined and largely arrhythmic nucleotide sequence.

c. Regularities

The mistakes of the present are made by trying to avoid the mistakes of the past. Though the danger of the premature imposition of regularities is well exemplified by the tetranucleotide hypothesis, it would be an error not to draw generalizing conclusions from experimentally established facts. The apodictic manner of the following enumeration is chosen to conserve space. Despite wide divergences in their nucleotide composition and sequence all deoxyribonucleic acids appear to be characterized

by the following regularities^{1, 21, 22}: (a) The sum of the purine nucleotides equals that of the pyrimidine nucleotides. (b) The molar ratio of adenine to thymine equals 1. (c) The molar ratio of guanine to cytosine (+ methylcytosine) equals 1. (d) The number of 6-amino groups (adenine + cytosine + methylcytosine) equals that of 6-keto groups (guanine + thymine). The last-mentioned regularity, and in some cases also the others, but with uracil taking the place of thymine, applies also to certain pentose nucleoproteins²³.

It is, in fact, because of these relationships that it is possible to state that, for instance, in the deoxypentose nucleic acids of wheat germ 5-methylcytosine substitutes for a part of the cytosine^{22, 24} or that in certain bacteriophage nucleic acids 5-hydroxymethylcytosine replaces cytosine entirely²⁵. It is, incidentally, remarkable that so far no purine satellites have been discovered.

But nature is inexorably approximate; and one could suspect that occasional deviations from fit are more significant than the fit itself. This may still turn out to be true of such giant molecules as the nucleic acids, but I do not believe it will. In any event, I should like to cite here a recent survey in which the results of all published analyses, carried out in several laboratories, of deoxypentose nucleic acid specimens from about fifty different genera were tabulated¹. The average values of the molar relationships that bring out the complementariness principles mentioned before were as follows.

Adenine to thymine	1.009
Guanine to cytosine (or its analogues)	1.001
Purines to pyrimidines	1.000
6-Amino to 6-keto groups	1.008

If it is considered that the nucleic acids varied very widely in their nucleotide composition (ratios of adenine + thymine to guanine + cytosine between 2 and 0.4), the agreement of the 1 : 1 relationships is most impressive, especially in view of the diversity of sources, preparations and procedures. No authentic case of an appreciable deviation has yet come to my attention.

d. Molecular structure

Deoxypentose nucleic acids are polynucleotide chains that are held together by phosphoric acid diester bridges between the deoxypentose molecules. Our knowledge of the sugar found in different preparations rests mostly on chromatographic evidence which has up to now yielded no indication of the presence of any but 2-deoxyribose¹. The position of the phosphate bridges has been studied in some detail²⁶ only in the deoxyribonucleic acid of calf thymus in which it appears to be 3',5'. No determinations of end groups have as yet been possible nor has the existence of a terminal phosphate group been demonstrated. Strictly speaking, it cannot yet be asserted that the deoxypentose nucleic acids occur as open and not as closed chains. It is, however, probably too early to worry about the tail before we can describe the animal.

The excellent X-ray photographs obtained at King's College^{27, 28} and the discovery of the general unity relationships mentioned before have formed the basis of an interesting structural proposal^{29, 30}. This model, which in the recent past has been unveiled repeatedly and sometimes very amusingly*, assumes a helical dyad in which two coiled strands are held together by a specific pairing of the nitrogenous constituents showing the unity relationships mentioned before. This conception has several attractive features; it makes good use of some of the experimentally established facts and is, on the whole, not incompatible with the X-ray³² and part of the chemical evidence. But between what is plausible and what is true in the natural sciences there lies an ocean which has seldom been crossed.

The invention of models is an essentially sterile undertaking unless two things are thereby accomplished: (a) a satisfying explanation of all known and sometimes seemingly contradictory

* "This pairing is likely to be so fundamental for biology that I cannot help wondering whether some day an enthusiastic scientist will christen his newborn twins Adenine and Thymine!" (Ref. 31). No doubt, this too will happen; but is "enthusiastic" the correct epithet?

facts; (b) an impulse to further and crucial experimentation. Therefore, before this structural hypothesis is elevated to the dignity of a natural law, a few remarks may be in order. The model provides a satisfactory explanation, though it probably is not the only possible one, of the unity relationships observed experimentally. Whether, in its detailed form, it does more than describe the structure of that portion of the processed nucleic acid preparation from which the diffraction patterns were obtained, remains, however, to be established. More work on genuine nucleoproteins, especially on compounds with proteins other than histone or protamine, *e.g.*, of the type found in tubercle bacilli⁸, is urgently required. The existence of two discrete and complementary chains has not been demonstrated; no deoxypentose nucleic acid preparation has shown noticeable deviations from the 1 : 1 ratios, not even for the ratio of purines to pyrimidines.

It is, however, the second part of the hypothesis, namely, the suggestion of a copying mechanism for the mutual reproduction of the two half-chains, that is more difficult to reconcile with some of what is known. It is assumed that a specific pairing is brought about by hydrogen-bond formation between the nitrogenous bases of the two chains in such a manner that a base having a keto group at the 6 position (guanine, thymine) can bond only with a base carrying a 6-amino group (adenine, cytosine), and that a purine always bonds with a pyrimidine, thus maintaining a constant distance between the two chains. It will be observed that the model is quite strict as regards certain restrictions governing the size of the partners (purine is linked to pyrimidine) and the position of the hydrogen bonds (purine-1 to pyrimidine-1; purine-6 to pyrimidine-6), but does not require the absence or presence of substituents in other positions. Thus, 5-methylcytosine or 5-hydroxymethylcytosine should be able to take the place of cytosine, and uracil and 5-methyluracil (thymine) should be similarly exchangeable, both in their positions on the "positive" or "negative" or during the automatic replication of the chains. That the latter replacement does not occur despite

the availability of uracil is amply demonstrated by the absence of uracil from deoxypentose nucleic acids.

As regards the 5-substituted cytosine derivatives, the situation is even more disturbing. One must surely assume that both cytosine and 5-methylcytosine are available to the cell during the formation of certain deoxypentose nucleic acids, since both pyrimidines are incorporated in what appears to be constant proportions. For instance, three different groups of workers, analyzing several independently prepared specimens of the deoxypentose nucleic acid of wheat germ, found the following figures for the 5-methylcytosine content (as moles per 100 gram-atoms of phosphorus)^{24, 33, 34}: 5.8, 6.0, 5.7. Similarly, the corresponding proportion of 5-methylcytosine in the nucleates from bovine tissues is around 1.3 (see the survey in Ref. 1). This constancy of incorporation must be due either to a specific selection operating during the reproduction of the nucleic acids or, much less likely, to an unexpectedly steady level in the supply of these pyrimidines or their precursors in the cell. In the latter case, the degree of incorporation of a pyrimidine and its 5-substituted analogue would be governed by the laws of probability and there could be no preferential accumulation of the satellite in certain portions of the polynucleotide chain or in one or the other of the fractions into which a total deoxypentose nucleic acid preparation can be separated. But the available evidence suggests that 5-methylcytosine does not replace cytosine at random and that the scheme is either incorrect or incomplete in some essential features. This evidence is twofold.

(1) During the fractionation of calf thymus deoxyribonucleic acid, which will be mentioned in the concluding section of this article, very significant divergences in the concentration of 5-methylcytosine in the several fractions were found (compare Table 17 which is taken from Ref. 5). Similar results, soon to be published, have been obtained with fractions of the deoxypentose nucleic acid of wheat germ³⁵.

(2) The action of pancreatic deoxyribonuclease on deoxyribonucleic acids results in the production of a considerable

TABLE 17

5-METHYLCYTOSINE DISTRIBUTION IN FRACTIONS OF CALF THYMUS
SODIUM DEOXYRIBONUCLEATE*

	<i>Fraction</i>			
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
NaCl molarity in extraction of nucleohistone gel	0.40	0.45	1.7	2.6
Corrected mean proportions:				
Thymine	24.2	24.0	29.0	28.7
Cytosine	23.9	23.4	19.8	21.3
5-Methylcytosine	1.9	2.7	1.2	0
Molar ratio:				
Thymine to 5-methylcytosine	12.7	8.9	24.2	—
Cytosine to 5-methylcytosine	12.6	8.7	16.5	—

* The mean proportions of each constituent have been corrected on the assumption that one-half of the nucleic acid P is contributed by the pyrimidine nucleotides. In one hydrolysis experiment with fraction 4 a minute amount of 5-methylcytosine (about 0.6 mole) was found.

proportion of dinucleotides which have recently been separated and analyzed³⁶. Unless the assumption is made that this enzyme distinguishes specifically between cytosine and its 5-methyl derivative, it is significant that the proportions of the various dinucleotides are far from what randomness of incorporation would predict. For instance, the frequency of methylcytosine being linked to guanine in a dinucleotide as compared with that of cytosine being so attached is 20 times as high in calf thymus nucleic acid, and 13 times as high in the wheat germ preparation, than would have been expected for non-selective incorporation. The unexpectedly high proportion of 5-methyldeoxycytidylic acid that is linked to deoxyguanylic acid^{36, 37} makes one, in fact, wonder whether it is not the adjoining nucleotide rather than the one opposite that directs the incorporation.

Two other findings could also be cited, though their detailed discussion would take us too far, namely, (a) the incorporation, from the medium, of 5-bromouracil or 5-iodouracil in the deoxypentose nucleic acids of certain deficient bacteria or of bacteriophages in replacement of part of the thymine^{38, 39}; and (b) the

complete replacement of cytosine by 5-hydroxymethylcytosine in certain phage nucleic acids²⁵. The latter synthesis, at any rate, may be presumed to occur in spite of the presence of the analogous pyrimidine, cytosine.

e. Is there more than one deoxypentose nucleic acid in a cell?

An answer to this question appeared very difficult for a long time. On the one hand, many different nucleic acid preparations from the same genus showed a remarkable constancy of composition. On the other hand, it appeared rather unlikely on general grounds that a cell, which is able to produce so many different proteins, should give rise to no more than one deoxypentose nucleic acid. But the task of separating a mixture of closely related polyampholytes of a very high molecular weight appeared hopeless; and one could ask whether a nucleic acid preparation comprising a very large number of individuals should not be treated as one compound. This is, indeed, a problem that leads directly into the epistemology of chemical identity; and many have found that it is an ungrateful task to spend one's life at the frontier of knowledge without the proper passport.

Later, a process for the fractionation of deoxypentose nucleic acids was discovered which is based on the fractional dissociation of a nucleoprotein, as discussed above (see also Fig. 2). The procedure, which led to the separation of many deoxypentose nucleate preparations into a series of fractions of divergent purine and pyrimidine contents, was first applied to calf thymus nucleohistone⁵ and later extended to the fractionation of other nucleic acids through their complexes with histone, globin⁶ or polylysine¹⁹. Subsequently, a similar arrangement was described in a preliminary form in which the fractionation was carried out by means of a histone-kieselguhr column⁷.

The gradual extraction of nucleohistone or of an artificially prepared protein nucleate with salt solutions of increasing strength in the presence of a denaturing phase, such as chloroform-pentanol, was found to yield a series of nucleic acid fractions with diminishing concentrations of guanine and cytosine and rising

Fig. 7. Composition of fractions (in moles per 100 gram-atoms of phosphorus) prepared by the fractional dissociation of artificial histone complexes with the indicated sodium deoxypentose nucleate preparations. Each block represents a fractionation run; the consecutively numbered bars within each block represent individual fractions, with the NaCl concentration of the solutions employed for extraction rising from left to right. The first column (T) indicates the average composition of the total deoxypentose nucleic acid preparations. The abscissa indicates the proportion of original nucleic acid P recovered in each fraction. (Taken from Chargaff¹.)

concentrations of adenine and thymine. What should, however, be stressed is that in all fractions the equimolarity of each pair of constituents and of total purines and pyrimidines is fully maintained. A selection of such fractionation experiments is shown in Fig. 7.

These findings, fully discussed in the publications cited above, suggest that the deoxypentose nucleic acid of a given cell is composed of a very large number of differently constituted individuals and that it is possible to achieve the resolution of this entire spectrum of structural gradations into several distinct "bands", in which all the regularities characteristic of the entire nucleic acid are maintained.

There is little doubt that these fractions should not be considered as individuals. If we are dealing with a continuous spectrum, as I have suggested, it may, in fact, be impossible to isolate what could be considered as a rigorously homogeneous specimen; it is not easy even to imagine a proper criterion for such a decision. Conceivably, no two nucleic acid molecules, within the same nucleus, are entirely identical. But here we clash against the definition of a macromolecule; and seeking refuge in statistics, the science of reconciling the parallels, we may conclude that it is hard to fight on a front where friend and enemy wear the same uniform.

REFERENCES

- ¹ E. CHARGAFF, in E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vol. I, Academic Press, New York, 1955, p. 307.
- ² R. MARKHAM AND J. D. SMITH, in H. NEURATH AND K. BAILEY (Eds.), *The Proteins*, Vol. 2, Academic Press, New York, 1954, p. 1.
- ³ B. MAGASANIK, in E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vol. I, Academic Press, New York, 1955, p. 373.
- ⁴ C. F. CRAMPTON, R. LIPSHITZ AND E. CHARGAFF, *J. Biol. Chem.*, 206 (1954) 499.
- ⁵ E. CHARGAFF, C. F. CRAMPTON AND R. LIPSHITZ, *Nature*, 172 (1953) 289.
- ⁶ C. F. CRAMPTON, R. LIPSHITZ AND E. CHARGAFF, *J. Biol. Chem.*, 211 (1954) 125.
- ⁷ G. L. BROWN AND M. WATSON, *Nature*, 172 (1953) 339.
- ⁸ E. CHARGAFF AND H. F. SAIDEL, *J. Biol. Chem.*, 177 (1949) 417.
- ⁹ A. E. MIRSKY AND A. W. POLLISTER, *Proc. Natl. Acad. Sci. U.S.*, 28 (1942) 344.
- ¹⁰ E. HAMMARSTEN, *Biochem. Z.*, 144 (1924) 383.
- ¹¹ J. A. V. BUTLER, P. F. DAVISON, D. W. F. JAMES AND K. V. SHOOTER, *Biochim. Biophys. Acta*, 13 (1954) 224.
- ¹² M. H. BERNSTEIN AND D. MAZIA, *Biochim. Biophys. Acta*, 10 (1953) 600; 11 (1953) 59.
- ¹³ A. W. POLLISTER AND A. E. MIRSKY, *J. Gen. Physiol.*, 30 (1946) 101.
- ¹⁴ B. MAGASANIK AND E. CHARGAFF, *Biochim. Biophys. Acta*, 7 (1951) 396.
- ¹⁵ K. FELIX, H. FISCHER, A. KREKELS AND R. MOHR, *Z. physiol. Chem., Hoppe-Seyler's*, 287 (1951) 224.
- ¹⁶ I. BANG, *Beitr. chem. Physiol. Pathol.*, 4 (1904) 331; 5 (1904) 317.
- ¹⁷ E. CHARGAFF, in *Some Conjugated Proteins*, New Brunswick, 1953, p. 36.
- ¹⁸ J. A. LUCY AND J. A. V. BUTLER, *Nature*, 174 (1954) 32.
- ¹⁹ P. SPITNIK, R. LIPSHITZ AND E. CHARGAFF, *J. Biol. Chem.*, 215 (1955) 765.
- ²⁰ C. F. CRAMPTON AND E. CHARGAFF, *J. Biol. Chem.*, 226 (1956) 157.
- ²¹ E. CHARGAFF, *Experientia*, 6 (1950) 201. [See Chapter 1 of this book.]
- ²² E. CHARGAFF, *Federation Proc.*, 10 (1951) 654. [See Chapter 2 of this book.]
- ²³ D. ELSON AND E. CHARGAFF, *Nature*, 173 (1954) 1037.
- ²⁴ G. R. WYATT, *Biochem. J.*, 48 (1951) 584.
- ²⁵ G. R. WYATT AND S. S. COHEN, *Biochem. J.*, 55 (1953) 774.
- ²⁶ D. M. BROWN AND A. R. TODD, in E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vol I, Academic Press, New York, 1955, p. 409.
- ²⁷ M. H. F. WILKINS, A. R. STOKES AND H. R. WILSON, *Nature*, 171 (1953) 738.

- 28 R. E. FRANKLIN AND R. G. GOSLING, *Nature*, 171 (1953) 740.
- 29 J. D. WATSON AND F. H. C. CRICK, *Nature*, 171 (1953) 737, 964; *Cold Spring Harbor Symposia Quant. Biol.*, 18 (1953) 123.
- 30 F. H. C. CRICK AND J. D. WATSON, *Proc. Roy. Soc. (London) A*, 223 (1954) 80.
- 31 F. H. C. CRICK, *Sci. American*, 191, No. 4 (1954) 54.
- 32 M. H. F. WILKINS, W. E. SEEDS, A. R. STOKES AND H. R. WILSON, *Nature*, 172 (1953) 759.
- 33 G. BRAWERMAN AND E. CHARGAFF, *J. Am. Chem. Soc.*, 73 (1951) 4052.
- 34 S. G. LALAND, W. G. OVEREND AND M. WEBB, *J. Chem. Soc.*, (1952) 3224.
- 35 R. LIPSHITZ AND E. CHARGAFF, *Biochim. Biophys. Acta*, 19 (1956) 256.
- 36 R. L. SINSHEIMER, *J. Biol. Chem.*, 208 (1954) 445.
- 37 J. D. SMITH AND R. MARKHAM, *Nature*, 170 (1952) 120.
- 38 D. B. DUNN AND J. D. SMITH, *Nature*, 174 (1954) 305.
- 39 S. ZAMENHOF AND G. GRIBOFF, *Nature*, 174 (1954) 306.

CHAPTER 4

*The Chemistry and Function of Nucleoproteins and Nucleic Acids**

1. INTRODUCTION

The progress in our knowledge of the chemistry and biology of the nucleic acids has been truly enormous in the last seven or eight years. As is customary in the natural sciences, the first conception and formulation of the problems soon was followed by the invention of the technical means required for their solution; and thereupon ensued an avalanche of papers of which some were important and more were trivial.

As there has appeared recently a very comprehensive treatise on the nucleic acids, which attempts to take stock of our present knowledge¹, there would be little sense in my engaging in a swimming contest with a leviathan. Hence, instead of citing much literature I shall refer, wherever possible, to the pertinent chapters of this book.

In the biological sciences, as in many others, we are faced at present with a silent struggle between those believing in first causes and looking for a teleological foundation of all occurrences and those who think that macrocosm and microcosm are governed by laws of probability. The clash between these two schools has found an epigrammatic expression in the anecdote relating a remark that Einstein is said to have made about the uncertainty principle. "But the Good Lord does not throw dice!"—a statement, I should like to add parenthetically, that itself severely restricts His omnipotence, since there is no earthly

* Reprinted with permission from *Rend. ist. lombardo sci. Pt. I*, 89 (1955) 101–115.

or heavenly reason why the Good Lord should not throw dice if He so desires.

Be that as it may, there is a strong tendency in all of us to establish some sort of hierarchy and to consider the body or the cell as operating under a chain of commands. Therefore, a bright graduate student will be able to tell us what some of the profoundest intellects in biology would have shuddered to contemplate, namely, "DNA makes RNA, and RNA makes protein". There is, of course, little ground for such a brash statement; but this is a field in which the winds of fashion blow strong, and they have been blowing from this direction for quite some time. The moment will soon have arrived when it will be possible to consider the correctness of this generalization.

In the following, I shall be able only to touch upon a few aspects of a very large theme, as they may occur to a speaker who attempts to give a comprehensive view, but has little time in which to do it.

2. NUCLEOPROTEINS

Our knowledge of the nucleoproteins has gone through a few, rather curious stages². At the very beginning, *i.e.*, in the times of Miescher, Hoppe-Seyler, Kossel, etc., there existed little doubt that the nucleic acids were, in the cell, attached to proteins and that these nucleoproteins were more than a mere mixture of the partners. The very scanty experimental means then available did not permit to go much further. These ancient masters, who combined a great respect for nature with a great reticence in its interpretation, were succeeded by much more impetuous and less cautious generations; and the nucleoproteins often were considered as obnoxious obstacles to the purification of the all-important proteins. It is surprising how little work has been done on nucleoproteins in the recent past. Even when the term nucleoprotein appears in the title of a publication, the latter will often be found to deal either with the nucleic acids themselves or with so-called model compounds.

The nucleoproteins are defined as conjugated proteins in which the nucleic acid functions as the prosthetic group and is linked to the protein by electrostatic attraction or by secondary valence forces, and sometimes by both. It is, of course, necessary to make a clear distinction between the deoxypentose and the pentose nucleoproteins. In the deoxypentose nucleoproteins³ the electrostatic contribution often is very pronounced, owing to the fact that the deoxyribonucleic acids frequently are found in association with proteins carrying strong positive charges, such as the protamines or histones. Such complexes with basic proteins are dissociated comparatively easily; and it is, for this reason, sometimes difficult to decide whether such a nucleoprotein really existed in the living cell or whether its isolation was the result of a fortuitous combination of solutes that happened to be present in the same extract.

In decided contrast to the deoxyribonucleohistones and protamines, most ribonucleoproteins^{4, 5} are very difficult to untie. This impedes the isolation of intact ribonucleic acids, since such preparations are so difficult to free of protein that, when this is accomplished, they can no longer be regarded as truly representative specimens. It would seem that electrostatic links are only of secondary importance in the architecture of most pentose nucleoproteins. It may be considered as a consequence of this stubborn resistance to dissociation on the part of the pentose nucleoproteins that nucleotide analyses carried out on the total conjugated nucleoprotein have, on the whole, given much more satisfactory results than those performed on isolated pentose nucleic acids⁶.

Of the representatives of the deoxynucleoprotein family it is the nucleohistone of calf thymus that has been studied more than any other and that is—probably quite wrongly—considered as the prototype of the deoxypentose nucleoproteins. Less work has been done on the nucleoprotamines, as they occur, for instance, in fish sperm; and very little is known about microbial deoxynucleoproteins. As to the pentose nucleoproteins, cytoplasmic preparations mostly derived from the microsome fraction⁷ and

complexes of bacterial origin can be studied. The most widely investigated nucleoproteins, of both the deoxy and pentose types, namely, the viruses and phages, probably represent more complex types of organization than are encountered in the conjugated proteins that I have mentioned above.

Two principal methods are used for the isolation of deoxyribonucleohistone and of similar complexes with protamine, namely, (a) extraction with strong salt solution; (b) extraction with solutions of very low ionic strength. The procedures using strong salt solutions are most widely used, though they would appear more objectionable, if the dissociation of the nucleoprotein and the formation of an artifact is to be avoided. These nucleoproteins are soluble at a very low electrolyte concentration. As the latter increases, the solubility first drops rapidly, reaching a minimum in isotonic solutions, and then rises gradually. Though variations occur, freshly prepared nucleohistone is, in general, soluble at NaCl concentrations higher than 0.7 or 0.8 *M*.

The course of dissociation of two nucleoproteins (calf thymus and wheat germ) at rising NaCl concentrations is illustrated in Fig. 2 (p. 45, Ref. 8) and is based on experiments of Crampton *et al.*⁹ and of Lipshitz and Chargaff¹⁰. The observation that a progressive separation between protein and nucleic acid takes place has made possible the fractionation of the latter, as will be mentioned below.

3. DEOXPENTOSE NUCLEIC ACIDS

The position of the phosphate bridges holding together the mononucleotides of a polynucleotide chain has been studied in detail only in calf thymus deoxyribonucleic acid; the available evidence points to this link going from the 5' position to the 3' position on the neighboring nucleoside¹¹. Whether this generalization holds for many, or all, deoxypentose nucleic acids cannot yet be stated. Wherever the sugar moiety could be identified, it proved to be 2-deoxyribose³; but here, again, no general statement can be made. As to the nitrogenous constituents, two

purines, adenine and guanine, and two pyrimidines, cytosine and thymine, occur in all deoxypentose nucleic acids, with the exception of certain *E. coli* phages in which cytosine is replaced by 5-hydroxymethylcytosine. A minor, fifth component is often encountered, viz., 5-methylcytosine. For the structure of the nucleosides and nucleotides a recent survey may be consulted¹².

The evidence accumulated thus far indicates that there exists a very large number of different nucleic acids, which can be distinguished by different proportions, and therefore by differences in the sequence, of the component nucleotides³. The two most extreme examples are the deoxyribonucleic acid from a sea urchin, in which the ratio of the sum of adenine and thymine to that of guanine and cytosine ("the dissymmetry ratio") is 1.9 ("AT type"), and, on the other hand, the deoxyribonucleic acid of the avian tubercle bacillus in which it is 0.4 ("GC type"). While species specificity of composition can often be demonstrated, no indications of organ specificity have as yet emerged.

The incorrect tetranucleotide hypothesis has been abandoned; and it must be concluded that the nucleic acids are complicated and intricate high polymers, devoid of a recognizable periodicity. Despite wide divergences in their composition all deoxyribonucleic acid preparations appear to be characterized by several regularities³. These are: (a) $A + G = C + T$; (b) $A = T$; (c) $G = C$; and, therefore, (d) $A + C = G + T^*$. Only the last regularity, with thymine being replaced by uracil — $A + C = G + U$ — seems to apply to the ribonucleic acids⁶; in other words, the molar sum of compounds carrying an amino group in 6 position equals that of compounds with a 6-keto group in both types of nucleic acid.

4. FRACTIONATION OF DEOXPENTOSE NUCLEIC ACIDS

Throughout all our work on the composition of nucleic acids we were beset by one question: does a nucleic acid preparation,

* Abbreviations: A, adenine; G, guanine; C, cytosine; T, thymine; U, uracil; or the corresponding nucleotides.

as it is isolated from a given cell, represent one molecular species, several or many? On the one hand, we had observed a remarkable constancy in composition, at any rate of the deoxypentose nucleic acids from the same species. On the other hand, it appeared unlikely on general biological grounds that a cell gave rise to only one kind of deoxypentose nucleic acid. But the task of separating a mixture of closely related polyampholytes of a very high molecular weight appeared hopeless. Often, when I discussed this problem with my colleagues, I was asked what could be done about it; and my answer was that, if we were dealing with three or four different nucleic acids, a separation may be feasible, but that with a very large number of individuals in the same preparation we might as well treat them as one compound. This is, indeed, a problem that leads us directly into very deep waters, namely, the meaning of chemical identity in macromolecules.

In any event, it occurred to us that a separation, even if partial, could perhaps be effected at the nucleoprotein stage. As I mentioned before, the fractional dissociation of nucleohistone by salt solutions of increasing concentration seemed to offer possibilities. This hope was realized¹³, and the findings were later extended in our laboratory^{10, 14, 15}, and confirmed in others^{16, 17}. (For a survey, compare Ref. 3). As an example, I am showing here the fractionation of a preparation of wheat germ nucleoprotein which is described in a paper that is in course of publication¹⁰. The data are assembled in Table 18. The typical picture can be seen: the fractions are characterized by progressively decreasing proportions of guanine, cytosine and 5-methylcytosine and progressively increasing quantities of adenine and thymine; but the unity relationships are maintained. There is noted, moreover, an uneven distribution of methylcytosine with respect to cytosine.

Numerous fractionation experiments with different nucleic acids, in many of which artificially prepared complexes with basic proteins^{10, 14} or polyelectrolytes¹⁵ were employed, have led us to conclude that a preparation of deoxyribonucleic acid may comprise a very large number of differently constituted individuals.

TABLE 18

COMPOSITION OF DNA FRACTIONS OBTAINED BY FRACTIONAL EXTRACTION OF WHEAT GERM NUCLEOPROTEIN GELS*

No.	Fractions		Moles per 100 gram-atoms P					Molar ratios				
	NaCl molarity	% of DNA in starting material	A	G	C	MC	T	$\frac{A+T}{G+C+MC}$	$\frac{Py}{Pu}$	$\frac{A}{T}$	$\frac{G}{C+MC}$	$\frac{C}{MC}$
I	0.4	8.6	26.2	25.5	16.8	6.4	25.1	1.05	1.07	1.04	1.10	2.62
II	0.45	4.7	25.5	25.9	17.9	6.6	24.1	0.98	1.06	1.06	1.06	2.71
III	0.5	13.7	26.7	24.1	16.9	6.1	26.3	1.12	1.03	1.01	1.05	2.77
IV	0.55	12.8	27.5	23.4	16.1	5.6	27.4	1.22	1.04	1.00	1.08	2.88
V	0.6	10.5	29.3	22.2	16.1	5.1	27.3	1.30	1.06	1.07	1.05	3.16
VI	0.65	12.3	29.0	22.7	16.0	4.9	27.4	1.29	1.07	1.06	1.08	3.27
VII	0.75	12.7	29.1	21.6	15.8	4.8	28.7	1.37	1.03	1.01	1.05	3.29
VIII	0.8	6.5	30.2	21.0	15.0	4.4	29.4	1.48	1.05	1.03	1.08	3.41
IX	0.9	5.9	30.0	21.2	15.5	4.1	29.3	1.45	1.05	1.02	1.08	3.78
Total DNA			28.1	21.8	16.8	5.9	27.4	1.25	1.00	1.03	0.96	2.85

* Taken from Lipshitz and Chargaff¹⁰. — Abbreviations: A, adenine; G, guanine; C, cytosine; MC, 5-methylcytosine; T, thymine; Pu, purines; Py, pyrimidines.

5. DIFFERENCES BETWEEN PENTOSE NUCLEIC ACIDS

The demonstration of clear-cut analytical differences in pentose nucleic acids has met with greater difficulties than was true of the deoxypentose nucleic acids. This was in part due to their greater lability, which rendered many of the results obtained on highly purified specimens doubtful, but it was also due to the fact that the spread in individual nucleotide proportions is by no means as great in PNA as in DNA. That there existed great differences between the ribonucleic acids of yeast and those of pancreas or liver had, however, been recognized early. The entire problem has been reviewed⁴. I shall limit myself here to quoting a few typical results from a recent paper in which total pentose nucleoprotein fractions were analyzed under conditions that could not have led to fractionation⁶. These results will be found in Table 19.

A comparison between pentose nucleic acids from different

TABLE 19
NUCLEOTIDE RATIOS IN PENTOSE NUCLEIC ACIDS*

Source	$\frac{Pu}{Py}$	$\frac{A + U}{G + C}$	$\frac{6-Am}{6-K}$
Ox liver	0.80	0.63	1.04
Rat kidney	0.96	0.66	1.00
Cytoplasm of rat liver and kidney	0.99	0.62	0.93
<i>Paracentrotus lividus</i> , eggs and embryos	1.08	0.77	0.99
Wheat germ	1.07	0.76	1.00
Yeast	1.00	1.12	0.93
<i>E. coli</i>	1.18	0.87	1.00
<i>M. phlei</i>	1.06	0.73	0.93

* Taken from Elson and Chargaff⁶, with the exception of the results on wheat germ PNA which are taken from Lipshitz and Chargaff¹⁰. — Abbreviations: A, adenylic acid; G, guanylic acid; C, cytidylic acid; U, uridylic acid; Pu, purine nucleotides; Py, pyrimidine nucleotides; 6-Am, A + C; 6-K, G + U.

species and tissues and from different locations within the cell has, moreover, demonstrated that while the differentiation between species and organs on analytical grounds meets with difficulty, differences between nuclear and cytoplasmic nucleic acids can be shown¹⁸.

6. REMARKS ON FUNCTIONS

The spectator who has followed the development of our knowledge of nucleic acids in the past few years cannot help feeling that the stage is set for a grand finale. Many an acrobat has announced the *salto mortale* to a gasping audience (hoping to change it into a *salto immortale*); and though often only a graceful pirouette could be seen, *tamen est laudanda voluntas*. But I believe that the deeper we get, the darker it becomes. It is true, the temptation is almost overwhelming to pour all into one pot: growth, and its change from benign to malignant; bacterial transformations; the transmission of hereditary properties; the multiplication of virus and phage; the synthesis of inducible enzymes; the differentiation of tissues; the formation of antibodies—everywhere an interplay between nucleic acids and proteins, a spinning wheel in which the thread makes the spindle and the spindle the thread. But if it is hard to recognize the similarities between what is ostensibly dissimilar, it is even more difficult to make out the differences between what appears so similar; and moderation still is among the tools of our science.

Brushing past, therefore, the heralds of immediate glory and convinced, as I am, that we are not obliged to solve the secret of life in 1955, I must conclude that the actual and established evidence on direct biological functions of the nucleic acids is rather meager. I should, furthermore, question whether a search for functions, at any rate in the narrowly anthropomorphic framework of a vulgar teleology, is what we need most at the moment.

The relevant information on the biological roles of the deoxy-pentose nucleic acids¹⁹ and the pentose nucleic acids⁵ has been

reviewed recently. I shall list here only briefly what to me appears most significant.

(a) The nucleic acids are ubiquitous and presumably indispensable constituents of living matter. Autarkic entities (cells and cell communities) appear to require the presence of both types of nucleic acid, parasitic entities (viruses, phages) the presence of one. It must be considered as highly important that the transfer of biological information can be effected, or aided, either by pentose nucleic acids (plant and, probably, animal viruses) or by deoxypentose nucleic acids (bacteriophages). Against the argument of omnipresence the objection could be raised that this hardly distinguishes the nucleic acids from other plastic components of living matter, *e.g.*, the proteins or the lipids, though it is doubtful whether the latter are required by parasites.

(b) The most direct evidence of an involvement of polynucleotides of the deoxy series in the genetic apparatus of the cell is due to the epochal work of Avery and his colleagues on the phenomena of bacterial transformation. There can be little doubt that our present understanding of the mechanisms that are brought into play during transformation is at a most rudimentary stage, but even less doubt that this discovery will be marked as an important date in the history of biology.

(c) The investigation of the mutagenic action spectrum of ultraviolet radiation has yielded curves that closely resemble the absorption spectrum of nucleic acids. Certain chemicals known to react with nucleic acids exhibit mutagenic effects¹⁹ and others known to be cytostatic or cytotoxic have been shown to be incorporated into nucleic acids²⁰.

(d) Boivin and his colleagues pointed out that the deoxypentose nucleic acid content of a diploid nucleus is constant for a given species and twice the amount found in the corresponding haploid nucleus²¹.

(e) Many lines of evidence suggest a role of pentose nucleic acids in protein synthesis and in morphogenesis⁵. The relationship has, however, remained purely formal and generic; no cor-

relation between specific sequences in the nucleic acid and specific synthetic capacities has been established.

(f) The fact that the nucleotide distribution in deoxypentose nucleic acids is characteristic of the species, and unchanged in different tissues of the same host, must be contrasted with the far-reaching changes undergone by the proteins.

But an old question must be asked before many of these arguments: What is a cause, what is a symptom? Are nucleic acids different because they come from different cells, or is this very difference one of the causes of biological specificity? We seem to have lost the ability to say "We don't know". But this is a privilege that I still wish to retain.

We may draw the following preliminary conclusions from the work on the chemistry of the nucleic acids. There exists a very large number of differently composed individuals which are distinguished by differences in the sequence of their component nucleotides; for we must remember that different composition signifies different sequence. No perceptible periodicity, nor a repeating unit, can be claimed; but the distribution does not appear to be random. In calf thymus deoxyribonucleic acid the purine nucleotides have a greater chance to be next to, or near, other purine nucleotides than next to pyrimidine nucleotides, and the converse must be true of the latter²². Similar conclusions can be drawn from our older work on ribonucleic acids²³, when account is taken of what is now known about the specificity of pancreatic ribonuclease²⁴.

Yet, deoxyribonucleic acids, whether the whole or the sub-fractions, all show the impressive regularities that I have mentioned before in Section 3; and most ribonucleic acids appear to exhibit at least one regularity (Table 19). There may be some that will ask whether there is any sense in chasing such regularities or even whether, in view of the mischief worked by the now defunct tetranucleotide hypothesis, a search for such norms is not outright deplorable. If this represented an esthetic quest for pre-established harmony or the music of the spheres, such objections would, perhaps, be justified. But quite apart from these

observations having grown out of experiments, we should not forget that regularities of this sort may be the best, or only, means of recognizing the existence of systems concerned with the preservation, or the transfer, of information: tasks that we should like to assign to the nucleic acids or, more probably, to the nucleoproteins. A man standing at the street entrance of a subway station may notice that the efflux of passengers, in contrast to their inflow, showed a very marked periodicity or regularity; he would rightly conclude that the trains operated according to a schedule, and this might be the only way in which he could reconstruct a timetable: the information was conveyed to the observer by the imposition of a regularity of which the passengers themselves could have been unaware.

In the speculations on the biological role of the nucleic acids the nucleoproteins usually are neglected. I believe this to be a mistake. There is little evidence that the nucleic acids occur in living matter in the free state. Though it is possible that the conjugated proteins in which nucleic acids are found represent artifacts resulting from fortuitous combination, this is not likely. One could, in fact, assume that if specific nucleotide sequences are the actual genetic determinants, they could give rise, in combination with proteins, to unique geometrical shapes, so that—through the association of a specific polynucleotide and a specific polypeptide—another dimension, as it were, is produced. I may refer to a discussion of some of these problems in which the inheritance of geometrical peculiarities was considered in analogy to problems in topology²⁵.

In any event, there can be little doubt that methods for the sequence analysis of nucleic acids are most urgently needed. We are still very far from it; we cannot yet say whether a nucleic acid molecule should be considered as the carrier of the information assigned to one gene or whether—and this appears more likely to me—one and the same molecule represented a string carrying multiple potential information, similar to the knot writing of ancient Peru, the quipu.

The copying mechanism suggested for deoxyribonucleic acid²⁶

provides an example of a system through which information is *preserved*. Our recent suggestion⁶, in which the regularity observed in ribonucleic acids, namely, the equality in the proportions of 6-amino and of 6-keto compounds (compare Table 19), is translated into a relationship with protein, whose regularly occurring peptide bonds could, through hydrogen bonding, impose this regularity upon two polynucleotide chains, is an example of a system through which information is *transferred*.

I believe, however, that while the nucleic acids, owing to the enormous number of possible sequential isomers, could contain enough codescripts to provide a universe with information, attempts to break the communications code of the cell are doomed to failure at the present very incomplete stage of our knowledge. Unless we are able to separate and to discriminate, we may find ourselves in the position of a man who taps all the wires of a telephone system simultaneously. It is, moreover, my impression that the present search for templates, in its extreme mechanomorphism, may well look childish in the future and that it may be wrong to consider the mechanisms through which inheritable characteristics are transmitted or those through which the cell repeats itself as proceeding in one direction only. With these words I have perhaps again tied the knot that I tried to open in the beginning, since it may be that the first causes themselves are subject to a random priority.

REFERENCES

- ¹ E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vols. I and II, Academic Press, New York, 1955.
- ² J. N. DAVIDSON AND E. CHARGAFF, in E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vol. I, Academic Press, New York, 1955, p. 1.
- ³ E. CHARGAFF, in E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vol. I, Academic Press, New York, 1955, p. 307.
- ⁴ B. MAGASANIK, in E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vol. I, Academic Press, New York, 1955, p. 373.
- ⁵ J. BRACHET, in E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vol. II, Academic Press, New York, 1955, p. 475.
- ⁶ D. ELSON AND E. CHARGAFF, *Biochim. Biophys. Acta*, 17 (1955) 367.
- ⁷ G. H. HOGEBOOM AND W. C. SCHNEIDER, in E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vol. II, Academic Press, New York, 1955, p. 199.
- ⁸ E. CHARGAFF, *Symposia Soc. Exptl. Biol.*, 9 (1955) 32. [See Chapter 3 of this book.]
- ⁹ C. F. CRAMPTON, R. LIPSHITZ AND E. CHARGAFF, *J. Biol. Chem.*, 206 (1954) 499.
- ¹⁰ R. LIPSHITZ AND E. CHARGAFF, *Biochim. Biophys. Acta*, 19 (1956) 256.
- ¹¹ D. M. BROWN AND A. R. TODD, in E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vol. I, Academic Press, New York, 1955, p. 409.
- ¹² J. BADDILEY, in E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vol. I, Academic Press, New York, 1955, p. 137.
- ¹³ E. CHARGAFF, C. F. CRAMPTON AND R. LIPSHITZ, *Nature*, 172 (1953) 289.
- ¹⁴ C. F. CRAMPTON, R. LIPSHITZ AND E. CHARGAFF, *J. Biol. Chem.*, 211 (1954) 125.
- ¹⁵ P. SPITNIK, R. LIPSHITZ AND E. CHARGAFF, *J. Biol. Chem.*, 215 (1955) 765.
- ¹⁶ G. L. BROWN AND M. WATSON, *Nature*, 172 (1953) 339.
- ¹⁷ J. A. LUCY AND J. A. V. BUTLER, *Nature*, 174 (1954) 32.
- ¹⁸ D. ELSON, L. W. TRENT AND E. CHARGAFF, *Biochim. Biophys. Acta*, 17 (1955) 362.
- ¹⁹ R. D. HOTCHKISS, in E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vol. II, Academic Press, New York, 1955, p. 435.
- ²⁰ G. B. BROWN AND P. M. ROLL, in E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vol. II, Academic Press, New York, 1955, p. 341.

- ²¹ R. VENDRELY, in E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vol. II, Academic Press, New York, 1955, p. 155.
- ²² C. TAMM, H. S. SHAPIRO, R. LIPSHITZ AND E. CHARGAFF, *J. Biol. Chem.*, 203 (1953) 673.
- ²³ B. MAGASANIK AND E. CHARGAFF, *Biochim. Biophys. Acta*, 7 (1951) 396.
- ²⁴ G. SCHMIDT, in E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vol. I, Academic Press, New York, 1955, p. 555.
- ²⁵ E. CHARGAFF, *Cold Spring Harbor Symposia Quant. Biol.*, 12 (1947) 28.
- ²⁶ J. D. WATSON AND F. H. C. CRICK, *Nature*, 171 (1953) 964.

*The Very Big and the Very Small:
Remarks on Conjugated Proteins**

It is an old experience in the natural sciences that what is poison for one generation often is honey for the succeeding one. Whether this indicates, in a given case, the dawn of a better era or an inurement by frequent exposure to sublethal doses of truth often cannot be decided without the perspective of centuries. There are important exceptions, but in general a scientific truth fades every 30 years, to be replaced by another equally evanescent. A well-designed and well-constructed chair lasts longer.

It is not very long ago that the extreme contempt for the amorphous and intractable, felt by generations of organic chemists (or at least by the second-rate specimens), has made room for the realization that there is little sense in treating living and growing tissue merely as the starting material for the isolation of well-behaved crystalline substances. In recent times, the respect for nature and its multiform manifestations (as such a very healthy sign) has, in fact, sometimes assumed exaggerated proportions; and it is occasionally necessary to point out that the living cell is not simply a macromolecule with a skin, or the bacteriophage a nucleoprotein with a tail. So-called model experiments often are carried to incredible lengths, prompting one to say that confusion superimposed on complexity may produce papers, but not results, and that a skunk dipped into chlorophyll is not yet an apple tree. The secret of the organization of the cell will not be found by a clever sleight of hand.

* Reprinted with permission from S. GRAFF (Ed.), *Essays in Biochemistry*, Wiley and Sons, New York, 1956, pp. 72-76.

It is, however, becoming clear that organization, as observed on the macroscopic and microscopic levels, must be matched, on the submicroscopic and molecular levels, by the existence of patterns in which the varied arrangement of a limited number of constituents serves to impress individuality and specificity on cells or cell communities. One could venture the opinion that biochemical evolution is accompanied, or indeed caused, by the formation of macromolecules of ever-increasing complexity composed of an ever-diminishing number of constituents. One mechanism by which this may be accomplished is that of conjugation. Many different conjugated proteins are being recognized, such as the nucleoproteins, the lipoproteins, the mucoproteins, or the chromoproteins; many enzymes that carry cofactors, distributed in specific positions on the protein molecule, belong to one or the other of these groups. But we encounter also lipopeptides, mucolipids, etc.; and many more unrecognized compounds of this type must daily be going down the drains of our laboratories than repose in the graveyards of our scientific journals. When I am sometimes told that biochemistry has "run out of good problems", I shudder and reply: "Biochemistry has not even begun!"

The general aspects of the problem of conjugation have rarely been formulated clearly. This is perhaps not surprising, for this class of substances has long found itself between two chairs, as it were: too big to be handled conveniently by the chemist; too small to be seen consistently by the morphologist. The chemist strove for the isolation of the smallest unit endowed with homogeneity; the biologist attempted the recognition of the simplest structure endowed with function. The prize went to the loudest prophet with monomaniac intent. Since scientists in our time have, on the whole, lost the ability to say "We don't know", it is usually the man with the premature explanations that brings home the bacon; and by the time he has been found out he will have devoured it.

Another obstacle to the recognition of the biological importance of the conjugated proteins may be seen in their being

usually considered apart from each other under the headings of their respective prosthetic groups. I am aware of only one instance, namely, in a symposium held some time ago (1953) at Rutgers University, in which an attempt was made to consider the conjugated proteins as a family of substances having more in common than the small print that they occupy in the current textbooks of biochemistry. But there can be little doubt that many, if not all, life processes take place on what may be considered the surfaces of conjugated proteins. It is, perhaps, not uninformative very briefly to compare two groups of conjugated proteins on which my laboratory has spent some effort, *viz.*, the nucleoproteins and the lipoproteins.

As concerns the nucleoproteins, a strict distinction must be made between the complexes containing deoxypentose nucleic acids and those in which a pentose nucleic acid acts as the prosthetic group. When the conjugated proteins associated with the same type of nucleic acid are compared, it will be noticed that their properties are governed by the type of protein they contain rather than by the composition of their nucleic acid moiety. This is undoubtedly due to the ability of proteins to differ much more from each other in their chemical and physical properties than do the nucleic acids. The deoxynucleoproteins are, in many cases, complexes whose protein moiety is represented by a protein of markedly basic properties, such as a histone or a protamine. Although it is inviting to consider such compounds as salts between the cationic protein and the anionic nucleic acid, this would be wrong. It is, for the moment at any rate, much safer to regard the nucleoprotamines and the nucleohistones as specifically conjoined complexes of a complicated and, thus far, little-understood geometry to which both electrostatic and secondary valence bonds contribute. What all these compounds, however, appear to have in common is that they are readily dissociated by high electrolyte concentrations and that under conditions permitting the removal of the protein moiety, by precipitation or denaturation, the nucleic acids are liberated.

Occasionally deoxynucleoproteins are encountered in micro-

organisms that are exceptional in resembling the ordinary pentose nucleoproteins rather than the nucleohistones. On the other hand, certain plant viruses represent exceptions to the behavior of the most commonly found pentose nucleoproteins. Of the latter, as they occur in the microsomes and the nucleoli, it may be said that they exhibit less of an electrostatic character than the deoxy-nucleoproteins. The bonds holding the ribonucleic acid to the protein are broken with much less ease; and one gains the impression that in this case the structure of the prosthetic group is inextricably associated with the structure of the entire conjugated protein. Whereas the dissociation of a nucleohistone may be compared to the removal of branches from the trunk of a tree, the separation of the ribonucleic acid and protein moieties of a ribonucleoprotein resembles much more the disentanglement of the warp and the woof of a fabric.

All nucleoproteins share, however, one important feature: they are combinations of two types of giant ampholytes, each of which can, and undoubtedly does, exhibit innumerable specificities as regards shape and constituent sequence. Their combination probably does add a new dimension; but each partner is, in itself, fully competent to maintain a specific pattern and to convey intricate information.

Of the lipoproteins, on the other hand, it could be said that it is only through the attachment of the monomeric lipids to a protein that a specific pattern of lipid arrangement becomes possible. It is not improbable that the future will show that certain lipids can exist in the cell in a polymerized form capable of exhibiting sequential specificity. But up to the present the lipids seem to be the only bulk components of tissues that must be assumed to exist principally in a monomeric form. If the establishment of specific arrangements is considered as an attribute of cellular organization, the formation of specific lipoproteins is one of the ways in which the lipids can take part in such specific patterns. Certain lipids probably are attached to the proteins by a combination of electrostatic and hydrogen bonds; others may occur as solutions in the lipid moieties of lipoproteins.

There is little evidence of the existence of covalent links between lipid and protein.

The two types of conjugated protein considered here very briefly are, to a certain extent, representative of conjugated proteins in general. The prosthetic group may be soluble or insoluble in water; it may be a monomer of comparatively simple structure, a mixture of monomers, or a macromolecule having itself a complicated structure and being capable of an intricate sequential specificity. As may have been gathered from what I said before, I consider the principle of conjugation as the main process through which nature makes big things bigger and small things big. Size, in compounds participating in the life of the cell, is probably not an accident. Moreover, such processes of pre-determined aggregation may be one of the ways in which what sometimes is stupidly referred to as the "assembly line" is realized in the living cell. The models of which we can conceive are probably no more than an absurd caricature of the synthetic mechanisms, a multiplicity of templates in space and templates in time, through which the organism maintains patterns of this high degree of complexity, unless we assume (and there is no reason for that) that what is duplicated is not really a duplicate.

Romantic deduction has done much harm in the sciences. But the use, the almost unpredictable use, of imagination is an essential element in the operations of the human mind. The injunction not to be astonished—*nil admirari*—is one of the most stupid legacies of antiquity. When we consider this ever-repeated giant throw of dice, this internally regulated cataract of reactions, sequences, and products, our first response must be a deep astonishment at a chaotic regularity which has thus far defied our understanding. Eternal surprise is the engine that drives the searching intellect. Let us hope that the coming generations will not have lost the ability to wonder about the many meanings of these palimpsests of nature.

CHAPTER 6

*Of Nucleic Acids and Nucleoproteins**

*That each from other differs, first confess;
Next, that he varies from himself no less.*

POPE | EPISTLE TO COBHAM | 19-20

1. INTRODUCTION

A speaker before a group such as I have the honor of addressing today has one of two choices: he can say very much about very little or very little about very much. There are, of course, two more choices possible; but the usual lecturer—grown fat in, if not from, the pursuit of science—will be too conceited to admit the one and too modest to consider the other. In any event, though it would be nice to be able to say very much about a very important matter, the title of my talk must have told you that nothing of the kind is contemplated. In fact, as I go on I may give the impression of asking questions rather than giving answers. I can only hope that I shall not find myself in the position of the Indian conjurer who after having fixed the traditional rope in the air climbed up to hang himself with it.

We live in a time that is drunk with experiments. Dubious results, dubiously paralleled, serve to establish so-called facts with a celerity that would make a monkey blush. A scientific Rip van Winkle, were he to return from only a few weeks' sleep, would find that in this short time ten people have described a "system that makes RNA" and ten, a "homogenate that makes DNA"; and twenty others have even "synthesized protein *in vitro*". But when he takes a closer look at the sort of RNA and DNA and protein that have been produced, he may decide to go back to sleep. Perhaps Aristotle would have felt at home, but would

* Reprinted with permission from *Harvey Lectures*, 1956–1957, 52 (1958) 57–73.

Plato, would Socrates? What Heraclitus would have said, I know. He would have said: "You do not make the same mess twice". For, really, our entire conception of the meaning of isolated cell constituents and of their action under the artificial conditions of customary experimentation is due for a revision. I, for one, would not be astonished if thirty or fifty years hence most of the terms and conceptions with which our science—perhaps then known as baroque biochemistry—has made us familiar would be deader than phlogiston is today.

2. THE SIGNIFICANCE OF ISOLATED CELL CONSTITUENTS

Now, let us ask a very silly question*. Can we be sure that the substances that we isolate from the cell exist in the living cell? I should say, the answer will have to be: They do and they do not. A simple monomer—a fatty acid, a sugar, etc.—which we find in a cell probably has not been produced *de novo* in the course of the devastating processes known as careful isolation. Even in such cases, to decide whether these substances occurred in the free state is a difficult matter. It becomes immensely more so when we come to high polymers often endowed, as in the case of proteins or nucleic acids, with a multiplicity of charges. I have little doubt that the physical description of the molecular arrangement of such complicated compounds, which must be drawn, quartered, pickled, or embalmed in order to be studied, defines the pleasing shape into which they can be put rather than the form in which they exist in life. For this reason, I look with esthetic pleasure but with the utmost diffidence and mental reserve on the various structural models and spiral contortions—beautiful examples of non-representational sculpture—that have been proposed for the nucleic acids. I should advise to wait and see. Models—in contrast to those who sat for Renoir—improve with age.

* "The 'silly question' is the first intimation of some totally novel development." L. PRICE, *Dialogues of Alfred North Whitehead* (Mentor Book Edition, New York, 1956, p. 145).

The organic chemist, the biochemist, the crystallographer, the biologist often look at the same object and they may even use the same term, but they mean different things. Our natural sciences, which once could find room in a single brain, have grown diffuse and multilingual. We are faced with a real problem of translation: from the molecular to the submicroscopic level; and then, in turn, to microscopic structures, morphological units, organelles, cells and cell communities. Where and when shall we find the dictionary that could help us in these translations? I should say, probably not before biochemistry has earned the first half of its name. It is, of course, not surprising that such a term as "nucleic acid" does not carry the same connotation in the several disciplines. A sculptor and a tailor naturally have different ideas of what makes a man: one looks at the torso, the other at the pants.

Nevertheless, we must be content with avoiding the avoidable and use the means that our science has given us. It is time for me to speak of nucleic acids.

3. TWO TYPES OF NUCLEIC ACID

You all know that the convenient slang terms DNA and RNA stand, respectively, for deoxyribonucleic and ribonucleic acids*. If this were a more strictly chemical talk I would point out that the sugars have been identified in only a few cases and that a more neutral designation, such as deoxypentose and pentose nucleic acids, would be safer. But tonight we may disregard that, as the nature of the sugars occurring in the nucleic acid specimens of which I am going to speak has in almost all cases been verified.

When about ten or eleven years ago my colleagues, first of all

* The following abbreviations will be used: DNA, deoxypentose nucleic acid(s); PNA, pentose nucleic acid(s); A, adenine; G, guanine; C, cytidine; MC, 5-methylcytosine; T, thymine; U, uracil (or the corresponding nucleotides); Pu, purines; Py, pyrimidines; 6-Am, adenine + cytosine; 6-K, guanine + uracil.

Dr. E. Vischer and Mrs. C. Green, and I embarked on the study of the chemistry of nucleic acids, the main known facts were based principally on the fundamental work of Miescher, Kossel, Levene, Steudel, Feulgen, Thannhauser, Hammarsten, Jorpes, Caspersson, Brachet, Bawden, Pirie, Stanley, Avery, and their collaborators^{1,2}. At that period, "the present stage of our knowledge of nucleic acid chemistry could perhaps be compared to that in which protein chemistry found itself at the beginning of this century"³. The composition of only two nucleic acid preparations, the deoxyribonucleic acid of calf thymus and the ribonucleic acid of yeast, was more or less completely known, and this only qualitatively. Although nucleic acids, at any rate calf thymus DNA and the pentose nucleic acids of plant viruses, were known to exist as high polymers, although viruses had been recognized as nucleoproteins, although the principle active in bacterial transformation had been identified as deoxypentose nucleic acid—and I could go on listing a few more disconcerting facts—it was almost generally assumed that nucleic acids were relatively simple compounds composed of a series of "tetranucleotides". As no methods were available to test this structural hypothesis, this simplification contributed considerably to the peace of mind of the chemist; and it continues to have this soothing effect on many investigators even now.

We approached the study of nucleic acid chemistry in the belief that the nucleic acids were complicated macromolecules, comparable to the proteins in their intricate and specific structure⁴. It was evident that for a discussion of the chemical specificity of the nucleic acids to become possible, one question had to be answered first: How much of what do they contain? This required, first of all, the elaboration of precise micromethods for both the qualitative and quantitative determination of all nucleic acid components—procedures that because of the strong ultra-violet absorption of the purines and pyrimidines were based on filter paper chromatography and spectrophotometry in the ultra-violet⁵⁻⁸.

The principal nitrogenous constituents of the nucleic acids are

adenine, guanine, cytosine, and thymine in DNA, and the first three and uracil in RNA. DNA appears somewhat more varied in the array of its constituents than RNA; satellites such as 5-methylcytosine or 6-*N*-methyladenine are found occasionally, and in the even coliphages 5-hydroxymethylcytosine occurs instead of cytosine. What should be borne in mind is that, in general, there are two purines and two pyrimidines and that one member of each of these two groups carries, in its 6-position, an amino group, the other a hydroxyl or, tautomerically, a keto group.

4. DEOXYPENTOSE NUCLEIC ACIDS

a. Composition and regularities

As early as in 1949, our studies had led to the conclusion that there existed a very large number of different deoxyribonucleic acids: different as regards the proportions, and therefore the sequence, of their component nucleotides, but of a composition characteristic of the species, though not of the tissue, from which they were derived^{4, 9, 10}. These conclusions have received ample confirmation since that time¹¹⁻¹³. In Table 20, I have selected a few analyses from our own work^{10, 13-18} in which DNA specimens from various sources, easily seen to be different in composition, are listed. The examples have been so chosen as to comprise the entire spread in divergences found so far. The dissymmetry ratio, *i.e.*, the ratio of the sum of adenine + thymine to that of guanine + cytosine + methylcytosine, ranges from approximately 1.9 to 0.4. We have referred¹⁹ to deoxypentose nucleic acids with a ratio above 1 as the AT type, to those with a ratio below 1 as the GC type; the latter appears to occur only in several micro-organisms. *Escherichia coli* belongs to an intermediate class with almost equal quantities of all components.

It is, of course, not always possible to distinguish between nucleic acids of very different origin through a study of their total composition. Two examples of analytically indistinguishable preparations^{14, 20} are listed in Table 21. In such instances, it is necessary to resort to other methods of differentiation; I shall

refer to them later. But one thing should be stressed here. Some people are bothered by such similarities as I have shown in Table 21, and I have heard it said, and also read it in one of the reviews of the recent treatise on nucleic acids¹, that there is no convincing

TABLE 20
COMPOSITION OF SEVERAL TOTAL DEOXYRIBONUCLEIC ACIDS

No.	Origin	Moles per 100 gram-atoms P					Dissym- metry ratio
		A	G	C	MC	T	$(A + T)/(G + C + MC)$
1	<i>Paracentrotus lividus</i>	32.8	17.7	16.2	1.1	32.1	1.85
2	Man	30.4	19.6	19.2	0.7	30.1	1.53
3	Ox	29.0	21.2	19.9	1.3	28.7	1.36
4	Wheat germ	28.1	21.8	16.8	5.9	27.4	1.25
5	<i>E. coli</i> (K-12)	26.0	24.9	25.2	—	23.9	1.00
6	Avian tubercle bacillus	15.1	34.9	35.4	—	14.6	0.42

evidence that the deoxypentose nucleic acids of mammalian tissues differ from one species to another. I should, however, like to submit an entirely opposite proposition, namely, that once it is shown that there exist different nucleic acids (Table 20), it is safer to assume that all nucleic acids are species-specific. For we

TABLE 21
EXAMPLE OF TWO DEOXYRIBONUCLEIC ACID SPECIMENS
INDISTINGUISHABLE ANALYTICALLY

Origin	Moles per 100 gram-atoms P				Dissymmetry ratio
	A	G	C	T	$(A + T)/(G + C)$
Sheep liver	29.3	20.7	20.8	29.2	1.41
Salmon sperm	29.7	20.8	20.4	29.1	1.43

must remember that the minimum estimate⁴ of the number of possible permutational isomers in which a substance of the composition of calf thymus DNA could exist is phantastically high, namely 10^{1500} . Actually, this number will be even much larger, as a molecular weight of 750,000 for DNA, used in these computations, is too low. How then expect two different organisms that cannot make identical proteins to be able to duplicate this enormous throw of dice?

If we take it for granted that there exist a very large number of different nucleic acids—a number presumably much larger than can be shown merely by analysis—we may go on to ask: Are there any principles that unify what appears so disparate? No simple or recognizable repeating unit can be made out in these giant polynucleotide chains; they do not seem to possess a perceptible periodicity. But we became aware very early in our work⁴ of several remarkable regularities apparently characteristic of all deoxypentose nucleic acids. They are as follows: (a) The sum of the purine nucleotides equals that of the pyrimidine nucleotides. (b) The molar proportion of adenine equals that of thymine. (c) The molar proportion of guanine equals that of cytosine (+ methylcytosine). (d) The number of 6-amino groups (adenine + cytosine + methylcytosine) equals that of 6-keto groups (guanine + thymine). These regularities may be considered as well established. (See Table XVII in Ref. 13.) Polynucleotide fragments produced by the enzymic degradation of DNA do not show these regularities²¹, a finding that will have to be taken into account should well substantiated deviations from these pairing principles be discovered.

b. Fractionation

Having plunged into a whirlpool, though decorated with harmonious surface ripples, we proceeded to make it even more chaotic. We asked a question. Granted, we said, that the DNA of a given species represents a completely arrhythmic, highly polymeric nucleotide chain, should we consider it as strictly homogeneous or as consisting of a bundle of different chains, dif-

fering from each other in the patterns of nucleotide sequence? I asked this question in our first paper on the chemistry of DNA⁹; but for several years I was resigned to there not being an answer, for there is nothing more difficult at the present stage of our knowledge than the fractionation of a family of closely related polyampholytes of a very high molecular weight. Fortunately, I had started—first together with Miss Lipshitz and later also with Dr. Crampton—on a study of calf thymus nucleohistone: its chemical properties and its behavior on dissociation in strong salt solutions²². We found that the extent of splitting of the conjugated protein into nucleic acid and protein was a function of the electrolyte concentration. When the DNA portions removed by fractional dissociation were analyzed, they were found to exhibit a curiously graded composition²³. The stepwise extraction of nucleohistone or of artificially prepared nucleates of proteins^{17, 18, 24} or polylysine²⁵ with salt solutions of increasing strength under conditions favorable for the denaturation of the protein moiety yielded a series of nucleic acid fractions with diminishing concentrations of guanine and cytosine and rising concentrations of adenine and thymine. The equimolarity of each pair of constituents and of total purines and pyrimidines was, however, fully maintained in all fractions¹³. The weighted average composition that could be computed from all fractions corresponded very closely to that found for total DNA preparations of the same species.

One may conclude that the DNA of a cell comprises an entire spectrum of differently constituted individuals; a spectrum, however, with “bands” of different position and intensity in different species. It is not inconceivable that no two nucleic acid molecules, within the same nucleus, are entirely identical²⁶: a prospect that would seem to condemn us to forced statistics for life.

c. Structure and sequence

The existing evidence makes it appear likely that the nucleic acids are chains in which nucleosides are linked to each other by phosphoric acid diester bridges between the 5'- and 3'-positions²⁷.

The "backbone", therefore, is represented by a regularly constructed polydeoxyribophosphate; and it is in the "superstructure" of purines and pyrimidines, linked glycosidically to 1' of the sugar, that the differences between nucleic acids must reside (Fig. 8). An alphabet of four or five letters may look meager, but if the words which these four letters spell out are 25,000-letter words, the result may be oppressively informative. Even if we had a rigorously homogeneous nucleic acid and procedures for its stepwise and orderly dismemberment, it would be hopeless to undertake the decoding of its nucleotide sequence. By the time the cryptographer had completed his task and written down the entire arrangement, evolution would probably have overtaken him and he could start over again.

A strict sequence analysis is, therefore, not only unattractive, but probably impossible, at any rate in DNA; an assay of the distribution density of individual components, *i.e.*, a survey of tendencies in nucleotide sequence, could, however, be considered. By survey I mean an attempt at crudely mapping the order in which the mononucleotides are aligned in a given nucleic acid chain. If this alignment does not take place at random, preferential arrangements discovered in the various preparations may furnish a means for their differentiation in addition to what total analysis may reveal. We have attempted to achieve this in three independent lines of investigation, based on the vertical or the horizontal degradation of the nucleic acids (p. 338 of Ref. 13) or on both.

Fig. 8. Schematic representation of a DNA segment. The purines are represented by black, the pyrimidines by white, symbols. There is an equal number of black and white squares and of black and white triangles. The dissymmetry ratio of squares to triangles is 1.5.

These are the three lines of attack; but I shall deal only with the third in some detail. (a) The action of crystalline pancreatic deoxyribonuclease on DNA shows a characteristic trend in the composition of small fragments formed at different periods; in addition, it brings about the formation of larger fragments exhibiting a distortion in the original internucleotide ratios^{21, 28, 29}. I believe it was through these studies that the arrhythmic character of nucleotide sequence in DNA was first recognized. (b) A partial horizontal degradation of deoxypentose nucleic acids leads to the class of compounds designated as apurinic acids, first studied with Drs. Tamm and Hodes³⁰. Their formation is based on the great lability of the glycosidic link of the purines. Apurinic acids may be considered as large polynucleotides, of at least 60 sugar phosphates and nucleotides in a row³¹, that have been deprived of their purines without distortion of the original interpyrimidine ratios*. Such products are much more labile to chemical, though not to enzymic³², attack than intact DNA; and the study of their degradation by alkali led to the conclusion that calf thymus DNA is rich in chains in which tracts of pyrimidine nucleotides alternate with stretches in which purine nucleotides predominate³³. Similar conclusions could later be drawn with respect to other deoxyribonucleic acids¹⁷. (c) The third approach to the problem of nucleotide arrangement undertaken in collaboration with Dr. Shapiro³⁴ (and in part unpublished) has to do with the production of pyrimidine nucleoside diphosphates. It was known since the early work of Levene and Thannhauser, and has also been confirmed more recently³⁵, that the acid treatment of DNA leads to the formation of some 3',5'-diphosphates of deoxycytidine and thymidine. The mechanism of this reaction and its significance were, however, left in doubt. Kinetic studies on the breakdown of small oligonucleotides prompt the conclusion that the first release of nucleoside diphosphates (30 min at 100° and 0.1 M H₂SO₄) is due to the liberation of those

* If in Fig. 8 the black symbols are taken to represent purines, the apurinic acid yielded by this sequence would contain deoxyribophosphate units in positions 1, 3, 8, 9, and 10.

pyrimidine derivatives that were flanked on both sides by purine nucleotides, whereas more prolonged treatment affects polypyrimidine stretches, with the glycoside fission of cytidine more readily brought about than that of thymidine. It was possible to develop methods for the quantitative estimation of the diphosphates and other small fragments of degradation. Table 22 gives a selection of some of the results. It is clear that this sort of study may be a powerful tool in bringing out variations in the nucleotide arrangement of different nucleic acids; it is possible to distinguish between specimens that cannot be distinguished by the total analysis of their constituents. It may be concluded that the nucleic acids listed in Table 22 represented chains consisting predominantly of polypurine and polypyrimidine units and that the arrangement of constituents was far from random.

5. PENTOSE NUCLEIC ACIDS AND NUCLEOPROTEINS

In glancing, more briefly, at the composition of various pentose nucleic acids³⁶ one has to state that nothing much could have been stated as long as isolated preparations were compared. The

TABLE 22
DEOXYCYTIDINE AND THYMIDINE 3',5'-DIPHOSPHATES FROM DNA*

	$O_{0.65}$	$M_{0.7}$	O_t	$O_{0.75}$	$M_{1.0}$	$Ar_{2.6}$
Molar ratio in DNA T/C	1.03	1.26	1.36	1.50	1.51	1.58
pTp** mole % of total T	23.5	12.8	15.4	15.9	19.9	16.0
pCp** mole % of total C	10.8	6.00	9.70	9.20	6.30	14.0
Molar ratio pTp/pCp	2.24	2.68	2.16	2.58	4.77	1.80

* 30 min, 100°, 0.1 M H₂SO₄; DNA preparations from ox = O, man = M, *Arbacia* = Ar. Subscripts: t = total DNA; the others indicate NaCl molarity at which fraction was obtained.

** pTp and pCp denote the 3',5'-diphosphates of thymidine and cytidine.

more rigorously purified the specimens were, the less constant was their composition, even when material from the same source was compared. In Table 23, a few examples have been chosen from our own work^{37, 38}. A complete compilation of analytical results has appeared recently³⁶. No generally applicable regularities in composition seem to impose themselves from such a survey.

TABLE 23

COMPOSITION OF SEVERAL PNA PREPARATIONS FROM YEAST AND LIVER*

<i>Source</i>	<i>No.</i>	<i>Ade- nylic acid</i>	<i>Gua- nylic acid</i>	<i>Cyti- dylic acid</i>	<i>Uri- dylic acid</i>	<i>Pu/Py</i>	<i>6-Am/6-K</i>
Yeast	1	30	30	19	21	1.5	1.0
	2	30	28	22	20	1.4	1.1
	3	26	28	22	24	1.2	0.9
	4	26	27	21	26	1.1	0.9
	5	21	30	23	26	1.0	0.8
Pig liver	1	16	38	33	13	1.2	1.0
	2	19	36	29	16	1.2	0.9

* In moles per 100 moles nucleotide.

It was only after Dr. Elson in our laboratory undertook the study of the ribonucleotide distribution in extracts of the *total* PNA without further purification³⁹⁻⁴¹ that one striking regularity emerged^{42, 43}. Whereas none of the other regularities listed before for DNA applied to PNA in a general fashion, one persisted with very few exceptions, namely, the equality of constituents having 6-amino groups (adenylic and cytidylic acids) and of those carrying 6-keto groups (guanylic and uridylic acids): 6-Am = 6-K. (Compare the frequency distribution of nucleotide ratios in 104 PNA samples⁴³ illustrated in Fig. 9.) I show some typical results in Table 24 compiled from several recent papers^{18, 40, 43, 44}.

6. THE MEANING OF REGULARITIES

I hope I have shown that there is one regularity that almost all nucleic acids, deoxypentose and pentose nucleic acids, appear to

have in common, namely, the presence in them of an equal number of nucleotides carrying an amino group in 6-position and of those having a 6-keto group. Clearly, if this contention is correct there must exist an agent imposing this regularity on the poly-

Fig. 9. Frequency distribution of various nucleotide ratios in 104 PNA analyses, based on the data of Elson and Chargaff⁴³. The symbols are explained in the footnote on page 84.

TABLE 24
NUCLEOTIDE RATIOS IN PENTOSE NUCLEOPROTEINS

Source	Pu/Py	(A + U)/(G + C)	6-Am/6-K
Ox liver	0.80	0.63	1.04
Rat kidney	0.96	0.66	1.00
Cytoplasm of rat liver and kidney	0.99	0.62	0.93
<i>Paracentrotus lividus</i> , eggs and embryos	1.08	0.77	0.99
Wheat germ	1.07	0.76	1.00
Yeast	1.00	1.12	0.93
<i>Escherichia coli</i>	1.18	0.87	1.00
<i>Mycobacterium phlei</i>	1.06	0.73	0.93
<i>Azotobacter vinelandii</i>	1.21	0.78	1.00

nucleotide chains. It must select, by attracting or by rejecting, those components that fit, out of the surplus of innumerable metabolites being offered in the course of the life of the cell. The screen, the sieve, the template: how many mechanomorphic contrivances could one not think of! I believe, however, that this is a riddle that will not be solved through allegory. It is quite possible that our biochemical thinking will have to acquire an entire new dimension before we shall stop talking complete nonsense. But I hope I shall not be accused of preaching metabiochemistry.

There will be many among us who, having learned from the often experienced defeat of easy or premature generalizations in biochemistry, are allergic to attempts to bring order to the seeming chaos of the living cell. They may rightly ask: Is there any sense in looking for regularities in composition, such as I have discussed before; to apply interior decoration to a continuum? They are, of course, correct in protesting against the dictatorial sway of half-baked hypotheses. But there will always be a time in the natural sciences—and it will always be too early—when a summary, a provisional and tentative summary, must be drawn up. As long as we realize that the experimental sciences operate under an unwritten statute of limitations, no harm will result.

In the case of the nucleic acids it is particularly important to ascertain whether any unifying principles governing their construction can be discerned. They are considered by many people to be the carriers of biological information. This rather ill-defined notion views the DNA as the holder of the majority vote in the transmission of hereditary properties and the RNA as the Executive Vice-President in charge of protein production. Be that as it may—and there may be some surprises around the corner—biological information must be preserved; it must be transferred; and, in the case of mutations, it must be changed. An incredible amount of male handicraft is being wasted on information models. But if there is any meaning to the concept of biological information, the existence of regularities of the sort pointed out by me before may furnish us the only chemical clue to systems through which information is preserved or—and this is more important—

transmitted. Compositional regularities in themselves are not sufficient for an explanation; but they will, I believe, have to be taken into account in any eventual solution of this problem. You may not be able to reconstruct an animal from its tracks; but you may succeed in identifying it through its footprints.

7. REMARKS ON NUCLEOPROTEINS

During my entire talk this evening, there has loomed in the background, without ever making a real appearance, the second half of my title, the nucleoproteins. Immanently, they must be present in all our thinking on nucleic acids, for there is little evidence that nucleic acids occur in the free state in the living cell. If they ever disrobe, it is only temporarily; for the most part they appear dressed decently in a coat of protein. Why not, therefore, consider the whole thing, the conjugated proteins; the more so since there is much evidence of the close connection between nucleic acid and protein synthesis? In reality, we are not ready for it despite the merry voices of innumerable biochemical boy scouts who have pitched their hasty tents all over the field.

But it is probably not an accident that nucleic acids and proteins are so closely associated in nature. It is hard to believe that the proteins, in all their variety, have only the function of an ornament or wrapping material. Even in the plant viruses, I do not find it easy to entertain the Cytherean vision of a nucleic acid resting in a merely protective shell of protein.

I have no time to go into all that and can only refer to a few recent discussions^{13, 45-47}. But I should like to repeat one suggestion⁴³, namely, that the regularity in PNA, and perhaps also in DNA, of which I have spoken, $6\text{-Am} = 6\text{-K}$, could be imposed by an extended polypeptide chain bonded to two polynucleotide chains so that each peptide carbonyl is linked by a hydrogen bond to the 6-amino group of adenine or cytosine and each peptide amino group to the 6-keto group of guanine or uracil. It is rather unlikely that such nucleoproteins will be found easily⁴⁸; the enforcement of this stoichiometry may be an essentially dynamic

process that cannot be tested by the static means of isolation. Other ways may have to be sought.

If I have touched so lightly upon so few aspects of an important problem, I must apologize; and I must apologize even more for dealing only with matters with which my colleagues and I have been directly concerned. But what sense would there have been in trying to outrun a mastodon in two volumes?¹ What such a treatment loses in validity, it may gain in intensity. For if there is one thing that I have learned during a life in science it is that even in the tiniest splinter there may mirror itself an entire world.

REFERENCES

- ¹ E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vols. I and II, Academic Press, New York, 1955.
- ² J. N. DAVIDSON AND E. CHARGAFF, in E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vol. I, Academic Press, New York, 1955, p. 1.
- ³ E. CHARGAFF AND E. VISCHER, *Ann. Rev. Biochem.*, 17 (1948) 201.
- ⁴ E. CHARGAFF, *Experientia*, 6 (1950) 201. [See Chapter 1 of this book.]
- ⁵ E. VISCHER AND E. CHARGAFF, *J. Biol. Chem.*, 168 (1947) 781.
- ⁶ E. VISCHER AND E. CHARGAFF, *J. Biol. Chem.*, 176 (1948) 703.
- ⁷ E. CHARGAFF, B. MAGASANIK, R. DONIGER AND E. VISCHER, *J. Am. Chem. Soc.*, 71 (1949) 1513.
- ⁸ B. MAGASANIK, E. VISCHER, R. DONIGER, D. ELSON AND E. CHARGAFF, *J. Biol. Chem.*, 186 (1950) 37.
- ⁹ E. CHARGAFF, E. VISCHER, R. DONIGER, C. GREEN AND F. MISANI, *J. Biol. Chem.*, 177 (1949) 405.
- ¹⁰ E. VISCHER, S. ZAMENHOF AND E. CHARGAFF, *J. Biol. Chem.*, 177 (1949) 429.
- ¹¹ E. CHARGAFF, *J. Cellular Comp. Physiol.*, 38, Suppl. 1 (1951) 41.
- ¹² E. CHARGAFF, *Federation Proc.*, 10 (1951) 654. [See Chapter 2 of this book.]
- ¹³ E. CHARGAFF, in E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vol. I, Academic Press, New York, 1955, p. 307.
- ¹⁴ E. CHARGAFF AND R. LIPSHITZ, *J. Am. Chem. Soc.*, 75 (1953) 3658.
- ¹⁵ E. CHARGAFF, R. LIPSHITZ AND C. GREEN, *J. Biol. Chem.*, 195 (1952) 155.
- ¹⁶ B. GANDELMAN, S. ZAMENHOF AND E. CHARGAFF, *Biochim. Biophys. Acta*, 9 (1952) 399.
- ¹⁷ M. E. HODES AND E. CHARGAFF, *Biochim. Biophys. Acta*, 22 (1956) 348.
- ¹⁸ R. LIPSHITZ AND E. CHARGAFF, *Biochim. Biophys. Acta*, 19 (1956) 256.
- ¹⁹ E. CHARGAFF, S. ZAMENHOF, G. BRAWERMAN AND L. KERIN, *J. Am. Chem. Soc.*, 72 (1950) 3825.
- ²⁰ E. CHARGAFF, R. LIPSHITZ, C. GREEN AND M. E. HODES, *J. Biol. Chem.*, 192 (1951) 223.
- ²¹ S. ZAMENHOF AND E. CHARGAFF, *J. Biol. Chem.*, 187 (1950) 1.
- ²² C. F. CRAMPTON, R. LIPSHITZ AND E. CHARGAFF, *J. Biol. Chem.*, 206 (1954) 499.
- ²³ E. CHARGAFF, C. F. CRAMPTON AND R. LIPSHITZ, *Nature*, 172 (1953) 289.
- ²⁴ C. F. CRAMPTON, R. LIPSHITZ AND E. CHARGAFF, *J. Biol. Chem.*, 211 (1954) 125.
- ²⁵ P. SPITNIK, R. LIPSHITZ AND E. CHARGAFF, *J. Biol. Chem.*, 215 (1955) 765.
- ²⁶ E. CHARGAFF, *Symposia Soc. Exptl. Biol.*, 9 (1955) 32. [See Chapter 3 of this book.]

- 27 D. M. BROWN AND A. R. TODD, in E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vol. I, Academic Press, New York, 1955, p. 409.
- 28 E. CHARGAFF AND H. S. SHAPIRO, *Exptl. Cell Research*, Suppl. 3 (1955) 64.
- 29 M. E. HODES AND E. CHARGAFF, *Biochim. Biophys. Acta*, 22 (1956) 361.
- 30 C. TAMM, M. E. HODES AND E. CHARGAFF, *J. Biol. Chem.*, 195 (1952) 49.
- 31 C. TAMM AND E. CHARGAFF, *J. Biol. Chem.*, 203 (1953) 689.
- 32 C. TAMM, H. S. SHAPIRO AND E. CHARGAFF, *J. Biol. Chem.*, 199 (1952) 313.
- 33 C. TAMM, H. S. SHAPIRO, R. LIPSHITZ AND E. CHARGAFF, *J. Biol. Chem.*, 203 (1953) 673.
- 34 H. S. SHAPIRO AND E. CHARGAFF, *Federation Proc.*, 15 (1956) 352.
- 35 C. A. DEKKER, A. M. MICHELSON AND A. R. TODD, *J. Chem. Soc.*, (1953) 947.
- 36 B. MAGASANIK, in E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vol. I, Academic Press, New York, 1955, p. 373.
- 37 E. CHARGAFF, B. MAGASANIK, E. VISCHER, C. GREEN, R. DONIGER AND D. ELSON, *J. Biol. Chem.*, 186 (1950) 51.
- 38 B. MAGASANIK AND E. CHARGAFF, *Biochim. Biophys. Acta*, 7 (1951) 396.
- 39 D. ELSON AND E. CHARGAFF, Phosphorus Metabolism, *Johns Hopkins Univ., McCollum-Pratt Inst., Contrib.*, 2 (1952) 329.
- 40 D. ELSON, T. GUSTAFSON AND E. CHARGAFF, *J. Biol. Chem.*, 209 (1954) 285.
- 41 D. ELSON, L. W. TRENT AND E. CHARGAFF, *Biochim. Biophys. Acta*, 17 (1955) 362.
- 42 D. ELSON AND E. CHARGAFF, *Nature*, 173 (1954) 1037.
- 43 D. ELSON AND E. CHARGAFF, *Biochim. Biophys. Acta*, 17 (1955) 367.
- 44 A. LOMBARD AND E. CHARGAFF, *Biochim. Biophys. Acta*, 20 (1956) 585.
- 45 J. BRACHET, in E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vol. II, Academic Press, New York, 1955, p. 475.
- 46 E. CHARGAFF, *Rend. ist. lombardo sci.*, 89 (1955) 101. [See Chapter 4 of this book.]
- 47 R. D. HOTCHKISS, in E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vol. II, Academic Press, New York, 1955, p. 435.
- 48 E. CHARGAFF, D. ELSON AND H. T. SHIGEURA, *Nature*, 178 (1956) 682.

*Nucleic Acids as Carriers of Biological Information**

1. IS THERE A HIERARCHY OF CELLULAR CONSTITUENTS?

Our time is probably the first in which mythology has penetrated to the molecular level. I read, for instance, in a recent article by a very distinguished biologist:

"... In the early phases of the molecular stage of evolution, only simple molecules were formed... Later more complex molecules, such as amino acids and perhaps simple peptides, were formed.

"In the more advanced phases of this period it is believed that there appeared a molecule with two entirely new properties: the ability systematically to direct the formation of copies of itself from an array of simpler building blocks, and the property of acquiring new chemical configurations without loss of ability to reproduce. These properties, self-duplication and mutation, are characteristic of all living systems and they may therefore be said to provide an objective basis for defining the living state.

"Evidence is accumulating that the nucleic acids of present-day organisms possess these two properties, and it is perhaps no longer useless to speculate that the first "living" molecule might have been a simple nucleic acid, perhaps protected by an associated simple protein..."

Thus, what started cosmically with beautiful and profound legends has come down to a so-called "macromolecule". If poetry has suffered, precision has not gained. For we may ask whether a model that merely provides for one cell constituent continually to make itself can teach us much about life and its origins. We may also ask whether the postulation of a hierarchy of cellular constituents, in which the nucleic acids are elevated to a patriarchal role in the creation of living matter, is justified. I believe

* Reprinted with permission from "*The Origin of Life on the Earth*", *I.U.B. Symposium Series*, vol. 1, Pergamon Press, London, 1959, p. 297.

there is not sufficient evidence for so singling out this particular class of substances.

We know that autarkic entities (cells and cell communities) require the presence of a very large number of different compounds, foremost among which are the ubiquitous plastic constituents, namely: (1) nucleic acids (both of the deoxypentose and pentose types, DNA and PNA); (2) proteins; (3) lipids; (4) polysaccharides. Parasitic systems (viruses, phages) apparently require the presence of nucleic acids (DNA or PNA) and protein. But such a generic statement probably is no more meaningful than if we were to say that all machines consist of iron, copper, nickel, etc. For neither the knowledge of quantity nor that of quality alone can satisfy our inquiry; and for an understanding of what is meant by organization an entirely new dimension will have to be added: a dimension of which barely the foundations are discernible at present. Moreover, to return to the example of the machine, it is not likely that we could learn much about the "origin of the automobile" from an inspection of the parts of a present-day car; nor could such an examination help us to decide whether there did not once exist an automobile made of glass, though we may conclude that, at any rate under present conditions, it would hardly have survived as the fittest means of locomotion.

In any event, even if we commit the oversimplifying fallacy of postulating a virus-like structure as the first "living" molecule, we are still left to deal with a nucleoprotein: a singular that conceals a multitude of possible structures. If scientific facts were subject to a vote, the majority opinion would probably place the nucleic acids at the top of the hierarchy and let the proteins follow; something like this: DNA $\rightarrow$ PNA $\rightarrow$ protein. But in my opinion it would be more honest to confess that we know very little indeed about these things and to say that the road to the future should not be uselessly cluttered up with shoddy, and often entirely baseless, hypotheses. It may, however, not be un-instructive to say a few words about the chemical connotations of the concept of biological information.

2. ON BIOLOGICAL INFORMATION CHEMICALLY CONVEYED

A biochemist, when asked to consider this problem, probably would first of all translate it into the concept of chemical specificity—with which I have dealt in more detail on several occasions^{1,2}—and think of a substance of an elevated molecular weight, built of a number of chemically different monomers and possessing chemical or physical features which are preserved without change within the species, but which serve to distinguish the particular substance from analogous ones produced by other species. Specific polysaccharides, specific proteins, specific nucleic acids are the type of compounds that will come to mind immediately; and the immunological specificity exhibited by many representatives of the first two groups perhaps is their most striking property. The recognition of the existence of specific nucleic acids required much more time; but their specificity, too, can now be taken to be an established fact on both chemical and biological grounds¹⁻⁴. It is, indeed, only in the fourth group of principal cell constituents, namely the lipids, that difficulties with respect to the occurrence of species specificity still are encountered. As I have pointed out recently⁵, one way in which the cell may be able to render lipids specific is in their arrangement as prosthetic groups of lipoproteins. But there exist indications that nervous tissue contains complex mucolipids of high molecular weight which could very well exhibit species specificity⁶.

A warning is, however, in order. We cannot yet answer the question: How identical is identity? In considering cell replication, we like to assume that the complex and specific cell constituents are reproduced so as to give entirely invariant structures, with the hundreds or even thousands of different amino acids or nucleotides always arranged in an unchanged sequence. Actually, very little is known; the extent of play that nature may allow itself cannot yet be gauged.

It is, in any event, safe to assume that the real specificity of a given cell resides in the nature of its plastic constituents rather

than in that of its catalytic components, which latter probably are a symptom rather than a cause. But there is a giant step from the establishment of sequential specificity in proteins, nucleic acids, polysaccharides to the formulation of the manner in which the "information" stored in them is not only preserved, but also transferred from one class of components to the other. The polysaccharides and lipids cannot at all be fitted into what little we know. As regards the nucleic acids and proteins, however, it would appear to me safer, instead of postulating a hierarchy, to envisage a triangular relationship, *e.g.*, as follows:

No arrows have been placed in this diagram, for their direction may conceivably be different in different systems; so that, for instance, in a plant virus the relation is between 2 and 3, in bacteriophage duplication and in bacterial transformations between 1 and 3, in autarkic systems between 1, 2 and 3. It is, moreover, quite conceivable that the transfer of information often can proceed in either direction and that the prevailing impression that it invariably flows from the nucleic acids to the proteins is mistaken. We may not yet have learned how to isolate suitable protein preparations; I should not be surprised if it were found eventually that both halves of a plant virus, the PNA and the protein, may be "infective". It was, therefore, not without purpose that I used the plural above in speaking of "life and its origins".

One could, however, ask: Is the cell really nothing but a system of ingenious stamping presses, stencilling its way from life to death? Is life itself only an intricate chain of templates and catalysts and products? My answer to these and many similar questions would be No; for I believe that our science has become too mechanomorphic, that we talk in metaphors in order to conceal our ignorance, and that there are categories in biochemis-

try for which we lack even a proper notation, let alone an idea of their outlines and dimensions.

Regardless, however, of whether we accept the currently fashionable template hypothesis, we must assume that there exist agents or complexes of agents in the cell that preside over the selection and specific arrangement of the constituents of all cell-specific polymers: proteins, nucleic acids, blood group substances, bacterial polysaccharide antigens, etc. These agents must be able both to preserve and to transfer whatever codes are stored in the constituent sequence and the specific physical shapes of these compounds. How these agents operate is not yet within our power to describe; but there exists some evidence that the nucleic acids, both DNA and PNA, are concerned in these operations. It lies within the scope of these remarks to inquire whether there are any outstanding features that are common to all nucleic acids, for this may be of value for our understanding of the chemical characteristics of such information systems. In dividing the rest of the discussion under the headings of invariability, diversity, and regularity, it will be apparent that the first two qualities probably are common to all specific cell constituents, whereas the third is a most unusual feature of the nucleic acids which apparently is not shared by the other plastic cell substances.

3. INVARIABILITY OF NUCLEIC ACIDS

This is a property that is, often by tenet rather than by experience, held to be common to all macromolecular cell components. How limited our knowledge is, I have already emphasized before; but it is, in general, assumed that a given protein or specific polysaccharide or nucleic acid is invariable in structure in a particular species. Although, as I shall mention below, there exists impressive evidence that the nucleic acids of a cell comprise an entire series of different individuals, the constancy, with respect to characteristic base composition, of *total* deoxypentose nucleic acids appears well established²; a constancy that seems to apply to preparations from all tissues of the same

host⁷ or from different variants of the same microbial species⁸.

As concerns pentose nucleic acids a decision appeared more difficult for some time owing to the unsatisfactory nature of isolated PNA preparations. When methods were developed permitting the analysis of total PNA without previous isolation, essentially in the form of nucleoprotein, it could, however, be shown that the nucleotide composition of the total pentose nucleic acid of a given cell also is nearly constant^{9,10}. In more recent experiments on the pentose nucleic acids of *Azotobacter vinelandii*¹¹ and of *E. coli*¹² this remarkable invariability of PNA could also be demonstrated, in the latter case even under a variety of conditions, for instance, the presence or absence of simultaneous protein synthesis and with and without the addition of an excess of a single nucleoside.

4. DIVERSITY OF NUCLEIC ACIDS

Long before detailed chemical analysis had become possible the existence of many different proteins and polysaccharides had been recognized, often through their antigenic or other biological properties, characteristic physical qualities or some outstanding chemical components. The specific enzymic properties of many proteins also were noted very early. Direct biological tests are not often applicable to nucleic acids. The activity of specific deoxypentose nucleic acids in bacterial transformations is well established. The demonstration of the exclusive role of nucleic acids in virus growth and enzyme induction is, in my opinion, as yet far from unequivocal. But on the whole there surely exists enough evidence to speak of a diversity of nucleic acids on biological grounds.

The chemical diversity of DNA, as shown by widely different proportions of the constituent nucleotides in different specimens, has, since the time of its first discussion^{1,13}, been demonstrated in very many instances². Of more recent date is the discovery that a total DNA preparation can be fractionated into a whole series of differently composed, but regularly graded, fractions¹⁴. The

distribution of these fractions in a given cell must, however, be constant within narrow limits, as shown by the invariability of total DNA mentioned above.

The variation in the composition of different PNA preparations is marked, though perhaps not quite as striking as in DNA¹⁰. In addition, there is some evidence of the existence of different pentose nucleic acids in the same cell; *e.g.*, the nuclear and the cytoplasmic PNA of rat liver differ in composition¹⁵. A reliable chemical fractionation of PNA has, however, not yet been achieved.

5. REGULARITY OF NUCLEIC ACIDS

As I have mentioned before, this is perhaps the most unusual property distinguishing the nucleic acids from other cell-specific high polymers. In saying this, I am not referring to the regular position of the phosphodiester bridges connecting the nucleosides in the nucleic acid chain: they are assumed to be generally 3',5'-bridges; for the regularity of the links holding the monomers together is common to most biological polymers. What is so peculiar is the remarkable balance between the several constituents noticed in all deoxypentose and in almost all pentose nucleic acids: a type of equipoise that I am not aware of ever having been encountered in other mixed polymers that do not contain simple repeating units.

There are several regularities of which three are characteristic of only DNA^{1,2}. They are: (1) The molar quantity of adenine equals that of thymine. (2) The molar quantity of guanine equals that of cytosine (+ methylcytosine). (3) The sum of the purine nucleotides equals that of the pyrimidine nucleotides. The fourth regularity, finally, applies to nearly all nucleic acids, DNA and PNA¹⁰; it is: (4) The molar sum of nucleotides carrying 6-amino groups (adenylic, cytidylic acids) equals that of nucleotides having 6-keto groups (guanylic, thymidylic or uridylic acids).

In considering these regularities one must take into account that the nucleic acids are devoid of any perceptible periodicity or of

repeating sub-units larger than a mononucleotide. They are complicated high polymers of a largely arrhythmic nucleotide sequence which, however, does not appear to be fortuitous¹⁶.

6. CONCLUDING REMARKS

It is inviting to assume that the special biological functions of the nucleic acids are reflected in those chemical features that distinguish them from other high-molecular cell components, namely, in the unusual regularities in nucleotide composition mentioned above. In which way these functions are exerted cannot yet be said; but I already have pointed out before that in my opinion the nucleoproteins rather than the separate moieties of the conjugated proteins will eventually be found to be the operative units. A detailed inquiry into the mechanism of this conjunction of protein and nucleic acid, let alone the construction of a reasonable model or the attempt at unravelling the information code of the cell, is premature; for no truly homogeneous nucleoprotein has yet been discovered of which it could be said that the particular nucleic acid molecule contained in it, with its unique nucleotide sequence, has given rise to the particular protein, with its equally unique amino acid sequence, or *vice versa*. In other words, we are still lacking the Rosetta stone of biochemistry.

It is not profitable to speculate on the direction in which the ancestral roles of the nucleic acids could be approached. But it can be predicted that further investigations along the following lines will be of value: (a) virus nucleoproteins; (b) mechanism of enzyme induction; (c) mechanism of antibody production; (d) role of priming substances in the enzymic synthesis of high polymers.

REFERENCES

- ¹ E. CHARGAFF, *Experientia*, 6 (1950) 201. [See Chapter 1 of this book.]
- ² E. CHARGAFF, in E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vol. I, Academic Press, New York, 1955, p. 307.
- ³ R. D. HOTCHKISS, in E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vol. II, Academic Press, New York, 1955, p. 435.
- ⁴ J. BRACHET, in E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vol. II, Academic Press, New York, 1955, p. 475.
- ⁵ E. CHARGAFF, in S. GRAFF (Ed.), *Essays in Biochemistry*, Wiley and Sons, New York, 1956, p. 72. [See Chapter 5 of this book.]
- ⁶ A. ROSENBERG AND E. CHARGAFF, *Biochim. Biophys. Acta*, 21 (1956) 588.
- ⁷ E. CHARGAFF AND R. LIPSHITZ, *J. Am. Chem. Soc.*, 75 (1953) 3658.
- ⁸ E. CHARGAFF, *Symposium sur le métabolisme microbien, Congr. intern. biochim., 2. Congr., Paris*, (1952) 41.
- ⁹ D. ELSON, T. GUSTAFSON AND E. CHARGAFF, *J. Biol. Chem.*, 209 (1954) 285.
- ¹⁰ D. ELSON AND E. CHARGAFF, *Biochim. Biophys. Acta*, 17 (1955) 367.
- ¹¹ A. LOMBARD AND E. CHARGAFF, *Biochim. Biophys. Acta*, 20 (1956) 585.
- ¹² A. LOMBARD AND E. CHARGAFF, *Biochim. Biophys. Acta*, 25 (1957) 549.
- ¹³ E. CHARGAFF, E. VISCHER, R. DONIGER, C. GREEN AND F. MISANI, *J. Biol. Chem.*, 177 (1949) 405.
- ¹⁴ E. CHARGAFF, C. F. CRAMPTON AND R. LIPSHITZ, *Nature*, 172 (1953) 289.
- ¹⁵ D. ELSON, L. W. TRENT AND E. CHARGAFF, *Biochim. Biophys. Acta*, 17 (1955) 362.
- ¹⁶ H. S. SHAPIRO AND E. CHARGAFF, *Biochim. Biophys. Acta*, 23 (1957) 451.

CHAPTER 8

*First Steps towards a Chemistry of Heredity**

1. INTRODUCTION

An observer of our biological sciences today sees dark figures moving over a bridge of glass. We are faced with an ever expanding universe of light and darkness. The greater the circle of understanding becomes, the greater is the circumference of surrounding ignorance. But as our sciences expand, their specific gravity appears to become lower. Out of swimmers we have all turned into floaters. A unified and consistent vision of nature has become impossible in our days, at any rate for working scientists. Ironically enough, the only universal scientists left are the publishers of scientific books or the writers of science fiction. Each science protects itself from its neighbors by a cordon of slogans and catchwords; and fashion dictates whether this year we are featuring enzymes or proteins or nucleic acids and whether we wear the molecules long or short. New journals are born every day by Caesarean section performed by skilful publishers; and as new disciplines are formed, so are new and mutually unintelligible languages: a Tower of Babel made of paper.

The feeling of discouragement before the seemingly chaotic wave of nature, before this self-duplicating cataract, is not new. Long before the unjustly famous "Ignorabimus" of Du Bois-Reymond, innumerable mournful voices could be heard, the most intense, the most heart-rending, being that of Pascal; for here is

* Reprinted with permission from *Intern. Congr. Biochem.*, 4th Congr., Vienna, Austria, 14 (1958) 21-35.

a Moses who must curse the Promised Land*. One later citation may stand for many. When "in the late evening of an eventful life" one of the last, and one of the greatest, of universal scientists, Alexander von Humboldt, published his *Kosmos*, he had sworn off the vitalism of his youth. He wrote in his introduction:

"...Die Mythen von imponderablen Stoffen und von eigenen Lebenskräften in jeglichem Organismus verwickeln und trüben die Ansicht der Natur. Unter so verschiedenartigen Bedingnissen und Formen des Erkennens bewegt sich träge die schwere Last unseres angehäuften und jetzt so schnell anwachsenden empirischen Wissens. Die grübelnde Vernunft versucht muthvoll und mit wechselndem Glücke die alten Formen zu zerbrechen, durch welche man den widerstrebenden Stoff, wie durch mechanische Constructionen und Sinnbilder, zu beherrschen gewohnt ist.

Wir sind noch weit von dem Zeitpunkte entfernt, wo es möglich sein könnte alle unsere sinnlichen Anschauungen zur Einheit des Naturbegriffs zu concentriren. Es darf zweifelhaft genannt werden, ob dieser Zeitpunkt je herannahen wird. Die Complication des Problems und die Unermesslichkeit des *Kosmos* vereiteln fast die Hoffnung dazu. Wenn uns aber auch das Ganze unerreichbar ist, so bleibt doch die theilweise Lösung des Problems, das Streben nach dem *Verstehen* der Welterrscheinungen, der höchste und ewige Zweck aller Naturforschung..."

A. v. Humboldt, *Kosmos*, 1 (1845) 67.

The continuous struggle between vitalists and their adversaries was built on a solid rock of ignorance of what they really were fighting about. The best scientific discussions often take place in a thick mental fog. There can be little doubt: the ever renewed waves of vitalism have been repelled successfully; but so many victories have left our science short of breath. We shall not be searching here for a quaintly monistic "vital force". But just as much philosophy was necessary to dispose of the "philosopher's stone", much life will still have to flow past us, before we begin to understand what Life is. For have we really come nearer to the solution of the problem if the age-old question "What is life?" is replaced by an inquiry into the meaning of the organization of the cell? And if we no longer accept the God of Laplace, that most International of all Business Machines, are we so much better off babbling of "feedbacks"? We laugh at La Mettrie's

* See Pascal on the "vanité des sciences" in *Pensées* (Brunschvicg, No. 67, 74, 144, 604).

“L’homme machine”; but “La cellule machine” could be a contemporary title. The old anthropomorphism* of the sciences has been replaced by a streamlined mechanomorphism, with some sort of “templates” made of stainless protoplasm. But we may ask: Was this change really beneficial?

“Lasciate ogni speranza” is, however, hardly a motto that would appeal to the munificence of research foundations and government agencies; and a moderate optimism, intimating more than it ever promises outright, is the fashion of the day. We shall not dare transgress it, though no immediate vistas of roseate finality will be painted.

The title I have chosen for my talk is at the same time bold and modest. It should indicate how great the goal is and how far we are from it. But, as I recently said on another occasion¹, “there will always be a time in the natural sciences—and it will always be too early—when a summary, a provisional and tentative summary, must be drawn up. As long as we realize that the experimental sciences operate under an unwritten statute of limitations, no harm will result”. I shall first consider in what manner the phenomena of heredity may be subject to chemical research and what could be said to be the chemical basis of cellular specificity.

2. CHEMICAL BASIS OF CELLULAR SPECIFICITY

Let us start with a dictionary definition. “HEREDITY—3. *Biol.* The property of organic beings, in virtue of which offspring inherit the nature and characteristics of parents and ancestors generally; the tendency of like to beget like. (Often spoken of as a law of nature.)” Thus the great *Oxford Dictionary* (Vol. V, p. 238); and it is interesting to note that the term is not very old, the first example of its use dating from 1863. It is the sum of all the characteristics of a given cell which could be defined as *cell specificity*. But even if we are agreed that life at one level must

* “Il ne faut pas juger la nature selon nous, mais selon elle.” Pascal, *Pensées* (Brunschvicg, No. 457).

be all chemical, just as on another plane it is all physical, the translation of the concept of specificity into the terms of our science remains extremely difficult.

If we compare a *Staphylococcus* with an *E. coli* cell or a sea urchin egg with a mammalian ovum, comparative biochemistry would probably tell us that many of their metabolic reactions, many of their enzyme systems, are quite similar. Yet, how different is the whole! It must be confessed that there does not yet exist a good and all-embracing biochemical explanation of specificity. If we look at a cat and a mouse, we know that they are different—and they themselves certainly know it even better. But when we start to analyze them, we shall find much that is similar and so very little that is different. It is almost as if we tried to get at the meaning of “time” by taking apart a clock. There can be little doubt that comparative biochemistry is barely in its beginning and that, if it has difficulty in defining the differences between genera, let alone species, it is hardly equipped to deal with the distinction of individuals². The biology³ and immunology of individuality⁴ are much further advanced. It would be nice if this were merely another way of saying that we shall have to descend into nature by a few more decimals. But I am not certain whether a radical change in the direction of our efforts will not be necessary.

In any event, a biochemist meditating on the chemical basis of cellular specificity will be led directly to the concept of *chemical specificity*. He will assume—and I believe correctly—that differences in the observable behavior of cells must rest on differences in their chemical make-up and in the relative location of the interlocking cellular elements within the meshwork of the cell. As concerns the plastic constituents of the cell (proteins, nucleic acids, polysaccharides, lipids), the existence of important specificity characteristics within the groups of proteins and polysaccharides had been revealed by immunological studies long before organic chemistry had advanced to a point where chemical differentiation became possible. It took even longer for the chemical distinction of nucleic acids of different origin to become

feasible⁵. The lipids had to wait longest. Whereas the existence of many different tissue lipoproteins had been recognized for quite some time⁶, it is only recently that the occurrence of specific lipid polymers is being discussed^{7, 8}.

Which types of compounds then, we could ask, are responsible for the specific character of a cell? The answer is obvious. If we assume that all chemical and physical features of the cell are preserved without change within the species, but serve to distinguish the particular cell from that of other species, we must say that it is the aggregate of all cellular constituents that defines chemical specificity. But this does not mean that all substances are equally decisive in maintaining hereditary continuity; some may be causes, others merely symptoms. There is much reason for the assumption that the directive influence on the maintenance of cell specificity must be sought within the class of highly polymerized plastic constituents, at any rate, the proteins and the nucleic acids.

3. ON THE ALPHABET OF THE CELL AND SOME LANGUAGE DIFFICULTIES

When a science approaches the frontiers of its knowledge, it seeks refuge in allegory or in analogy. The latter attempt—and I consider it preferable to the first—has enlisted the support of modern disciplines, such as cybernetics or information theory⁹. A subordinate analogy to the maintenance of cellular specificity, but one, perhaps, more easily grasped, is that of the manner in which communication is brought about through language. This should not be taken to mean that I consider the cell as a system of phonemes; for we have no right to introduce such concepts as intelligence or mind into the working of the cell. But the comparison with communication through words will, at least, serve to explain the meaning of a term much used in these days, namely, *biological information*. I have dealt with this problem before^{10, 11} and can be brief here. A very tentative definition of biological information could describe it as the aggregate of all

the signals—*summa cellulae*—that preside over the maintenance of the hereditary processes, ensuring the unaltered transfer of all specificity characteristics of the mother cell to the daughter cells. We could then ask, but without much hope of an immediate and decisive answer: What is the content of biological information and what are its carriers; how is it preserved and how is it transferred; and what are the agents of such information transfers? If we knew the answers to these questions, we should have made more than the first steps towards a chemistry of heredity. If, on the other hand, we continue the criminal nonsense of filling our world with ever increasing quantities of ionizing radiation, we may soon be able to speak of the last steps towards a chemistry of heredity.

Fig. 10.

As to the content of biological information, the answer will depend upon whether the assumption I made a few minutes ago proves correct, namely, that the nucleic acids and the proteins, especially the conjugated proteins¹², are the principal, or even the sole, agents directing the maintenance of cellular specificity. In statements of this kind, allowance must be made for the necessity that we shall look as stupid to our successors of 2058 as our predecessors of 100 years ago look to us. In other words, we may be able to run away from our past, but never from our future.

The other questions are even more difficult to answer, for they refer to what may be called the flow of biological information; and it is here that our real language difficulties begin. There must exist in the cell a system of extremely rigid controls that supervise

the inviolate reproduction of all relevant constituents; a system presumably indispensable for the cell to maintain itself in a state of specificity. But the search for the nature of these controls has been largely unsuccessful. It has, however, led to the construction of many, often pretty, models which lately, since they are now made of transparent plastic, have gained much in astral beauty.

Our only refuge can be in vagueness, for instance, by proposing some scheme of triangular relationships which may take the childish geometrical form shown in Fig. 10. This is, of course, only one of many possible schemes for the interrelation of the plastic cell constituents. No arrows have been placed quite on purpose¹¹; and some of the solid lines should probably be broken lines or omitted entirely. The central role of ribonucleic acid is based on what is now beginning to be known of the metabolic functions of certain conjugated ribonucleotides. One of the main difficulties in all such schemes is that we do not yet know whether the same agent is concerned with the selection and alignment of the monomers and with the activation of their linking into polymers.

4. BIOCHEMICAL SPACE

When we speak of the "flow" of biological information¹¹ or of the "movement" of a given precursor, such as an amino acid or phosphoric acid, into the various polymers^{13, 14} found in the several cellular elements, such as nuclei, mitochondria, microsomes, etc., we adhere subconsciously to a concept that, I believe, is seldom formulated properly, namely, that of biochemical space. It is probably a thankless job to establish oneself as the Pindar of negative attributes; but there are many things in nature that are better described by a No than by a Yes. The cell, we might say, is not a machine. It is not a mixture of solids or even crystals. It is not a system of solutions. Even in describing it as an organized community of interfaces, of boundaries, we approach only a corner of the truth.

What would be necessary for an understanding of biochemical

space? We are all familiar with the four coordinates that define the space-time continuum. It is curious that in biochemistry the time coordinate is the one most easily defined, since through the use of isotopic tracers the dating of a polymer often can be achieved. But we know very little of the *spatial* conditions under which the cell manufactures its constituents. It is even possible—and this should not be forgotten in all the talk about “templates”—that we may be dealing with templates in time rather than space and that as yet undefinable time sequences regulate the replication of cellular polymers.

What I mean in speaking of the movement of precursors can, perhaps, best be illustrated by an example relating to the formation of protein and ribonucleic acid in rat liver^{13, 14} which is shown in Table 25. It will be seen that in experiments of short

TABLE 25

UPTAKE *in vivo* OF RADIOACTIVE LEUCINE BY THE PROTEIN AND OF RADIOACTIVE SODIUM PHOSPHATE BY THE RIBONUCLEIC ACID IN CYTOPLASMIC FRACTIONS FROM RAT LIVER^a

Fraction	Relative specific activity ^b							
	Protein		Ribonucleic acid ^c					
	Minutes after injection							
	5	20	5	20	120	1440		
Cytoplasmic supernatant fluid	1	1	11	9	20	7	10	1
Microsomal ribonucleoprotein particles ^d	I	4	2					
	II	6	2	1	1	1	1	1

^a The values are adapted from recent data^{13, 14}.

^b In each column the fraction having the lowest specific activity was taken as 1.

^c The specific activities of the nucleic acid samples were computed through the summation of the weighted specific activities of the 2' + 3' nucleotides produced by alkaline hydrolysis.

^d This represents the microsome portion insoluble in deoxycholate; the protein fraction I is insoluble at pH 13, fraction II insoluble at pH 6 (Ref. 15).

duration the protein precursor appears in the microsomal proteins before it is found in the supernatant fraction, whereas the opposite is true of the ribonucleic acid precursor. I have condensed the data and rounded off the figures for the sake of simplicity. If, in addition to the supernatant fraction (Sp), the microsomal portion soluble in deoxycholate (DS) and the two distinguishable protein fractions of the deoxycholate-insoluble microsome portion (RNP I and II) are surveyed, the movement of leucine into the proteins can be described as follows: $\text{RNP II} \rightarrow \text{RNP I} \rightarrow \text{DS} \rightarrow \text{Sp}$. An inverse, but probably correlative, direction is found for the movement of inorganic phosphate into the ribonucleic acids, namely, $\text{Sp} \rightarrow \text{DS} \rightarrow \text{RNP}$. The importance of such findings will emerge more clearly if we consider that not only the proteins, but also the nucleic acids, of the different cell fractions may be presumed to differ with respect to composition and function.

5. ON THE CODE-SCRIPT OF BIOLOGICAL HIGH POLYMERS

If you permit me to take a bird's-eye view—even if it is the view of a very shortsighted bird—I could ask: What is the immediate contribution that chemistry can make to an understanding of heredity? (Please notice that both here and in the title of my talk I used the word chemistry, not biochemistry.) I believe, I have already indicated that in my opinion this contribution must be sought in the complex of problems that has to do with the preservation, the transfer and the flow of biological information, and also with what could be called its geometry. We must gain a clearer understanding of the manner in which the ability to send the signals that direct the replication of macromolecules is embedded in the very structure of the various high polymers of the cell. I have often said that I do not believe that there exists a hierarchy among the cellular constituents; so that we should not say that, for instance, the nucleic acids or the proteins are more important than the lipids or the polysaccharides. But we can say that the minimum of viable organization is represented by a nucleoprotein, as in the viruses, and the minimum of signaling

potential by a deoxypentose nucleic acid, as in the transforming principles. Neither the nucleoprotein nor the nucleic acid can, of course, stand alone; they require the background of complete cellular organization. But at least the first step is indicated to us: we must attempt to read the code of the high polymers, first of all of the nucleic acids and the proteins. That this cannot take the shape of a legerdemain, of clever prestidigitation, is obvious. Little will be gained if the products formed in an uncooked meat extract or in a so-called homogenate are termed biosynthetic. But the search for the chemical structure of the decisive polymers, for the alignment, the sequence of their monomeric components is open to us.

The sequential analysis of the proteins and the nucleic acids will in itself not be sufficient; the less so, since no nucleoprotein has so far been isolated of which it could be assumed that the protein and nucleic acid moieties of the conjugated protein exhibited a direct relationship as of cause and effect. "We are still lacking the Rosetta stone of biochemistry¹¹." But such studies may indicate the direction that our future efforts are to take. As concerns the proteins, they will be discussed at this congress by more competent voices than could be mine. But since it has come to be that my laboratory has probably been among the first to place the structure of the nucleic acids into the context of modern biochemistry, you will permit me to devote most of the rest of my talk to this subject.

As I mentioned before, there exists some evidence that the nucleic acids, both the deoxypentose and the pentose nucleic acids, are concerned in the operations that have to do with the selection and the specific arrangement of the constituents of many, if not all, cell-specific polymers, such as proteins, nucleic acids, and perhaps even blood group substances and bacterial and other polysaccharides, etc. What then, could we ask, are the outstanding features that are common to all nucleic acids, but absent from other specific cell substances? This consideration may be of immediate importance for an understanding of the chemical characteristics of the information systems to which I

have alluded before. In all our following discussions we must keep in mind that the nucleic acids are devoid of any perceptible periodicity or of recognizable repeating subunits. They are high polymers of a largely arrhythmic nucleotide sequence which is, however, not random.

Until our first results were published in 1948 and 1949 (Refs. 16, 17) the nucleic acids, within their two known sugar categories, were considered as analytically indistinguishable; the problems posed by their analysis were not even seriously discussed. Now, the great *diversity* in the composition of the nucleic acids is recognized generally. I limit myself here to two examples (Tables 20, p. 87, and 26). Although occasionally nucleic acids from widely distant cellular sources are encountered that cannot be distinguished by constituent analysis, it is safe to assume that different species never produce identical nucleic acids.

But do the same cells always produce the same nucleic acids; in other words, do the nucleic acids exhibit the property of

TABLE 26
NUCLEOTIDE RATIOS IN RIBONUCLEOPROTEINS*

Source	Pu/Py	(A + U)/ (G + C)	6-Am/ 6-K
Ox liver	0.80	0.63	1.04
Rat kidney	0.96	0.66	1.00
Cytoplasm of rat liver and kidney	0.99	0.62	0.93
Nuclei of rat liver	0.66	0.81	1.06
<i>Paracentrotus lividus</i> , eggs and embryos	1.08	0.77	0.99
Wheat germ	1.07	0.76	1.00
Yeast	1.00	1.12	0.93
<i>Mycobacterium tuberculosis</i> (bovine)	1.06	0.65	0.98
<i>Mycobacterium phlei</i>	1.06	0.73	0.93
<i>Azotobacter vinelandii</i>	1.21	0.78	1.00
<i>Escherichia coli</i>	1.27	0.83	1.06

* Taken from previous papers¹⁸⁻²². Abbreviations: A, adenylic acid; G, guanylic acid; C, cytidylic acid; U, uridylic acid; Pu, purine nucleotides; Py, pyrimidine nucleotides; 6-Am (6-amino nucleotides), A + C; 6-K (6-keto nucleotides), G + U.

invariability within the species? This is apparently the case with respect to the total deoxypentose nucleic acid of a species, and also with respect to the total pentose nucleic acid of unicellular organisms. Many examples can be found in the literature^{20, 21, 23}.

The two properties I have mentioned, namely, diversity in composition with regard to different species and invariability of composition within the same species, are not unique with the nucleic acids; they are probably shared by all highly polymerized cell constituents, although it may be possible to distinguish between a direct, gene-controlled specificity (nucleic acids, proteins) and a reflected specificity (polysaccharides, lipids) which latter is mediated through specific enzymes that are themselves subject to control by the genes. But the nucleic acids show, in addition, a third, most unusual characteristic which distinguishes them from other cell-specific high polymers. I am referring to the well-known *compositional regularities*^{5, 18}. These are, in the case of deoxyribonucleic acids: (a) the molar sum of the purine nucleotides equals that of the pyrimidine nucleotides; (b) the molar proportion of adenine equals that of thymine; (c) the molar proportion of guanine equals that of cytosine (+ methylcytosine); (d) there is the same number of 6-amino nucleotides (adenylic, cytidylic acids, etc.), as of 6-keto nucleotides (guanylic, thymidylic acids). For most of the ribonucleic acids only the last regularity holds: the molar sum of adenylic and cytidylic acids equals that of guanylic and uridylic acids.

This remarkable balance between the several nucleic acid constituents represents a type of equipoise that I do not believe has ever been encountered in other mixed polymers that do not contain simple repeating units. That some of the principal beneficiaries of these discoveries have acknowledged them with great reluctance is a phenomenon not unusual in the history of the sciences.

One of the greatest discoveries in biology that I have had the good fortune to witness, namely, the recognition by Avery* and

* In the two recent textbooks of genetics that I have among my books the name of Avery does not appear. "*Habent sua fata*" not only the "*libelli*" but also what they leave out.

his collaborators of a specific deoxypentose nucleic acid as the agent responsible for bacterial transformation²⁴, has placed the nucleic acids, or at any rate the deoxypentose nucleic acids, very near the gene of the geneticist. If then these nucleic acids are, indeed, the genetic determinants, or form part of these systems, one could ask in what manner their biological functions can be translated into chemical structure, for translation there must be. Two possibilities suggest themselves. (1) Each nucleic acid chain, containing, perhaps, 10,000 or more nucleotides in a row, carries many single determinants. But then each gene could hardly be represented by a simple oligonucleotide sequence, say, a tri- or tetranucleotide, since it is highly improbable statistically that such sequences could be unique. (2) Less implausibly, each gene is represented, or determined, by a separate nucleic acid which is a distinct chemical entity, the quite well established constancy of composition of the total deoxyribonucleic acid of a cell being merely the statistical consequence of the unchanging character of the cell. In this case, a genetic determinant, defined by a characteristic chemical structure, would be unique if no two nucleic acid molecules, within the same cell, were identical. I have mentioned this possibility of nonidentity before²⁵, and it actually is not such an unreasonable assumption: a haploid nucleus containing, let us say, $3 \cdot 10^{-12}$ g of deoxypentose nucleic acid²⁶ would carry 300,000 molecules of nucleic acid of molecular weight $6 \cdot 10^6$ and correspondingly fewer if a higher molecular weight is taken. The great difficulty in such speculations is to reconcile the position of people for whom the gene is like a state of mind with that of others for whom it is a substance. Who would ask himself how many memories he has of the same event?

It has, in fact, been possible, first in our laboratory²⁷, to fractionate total deoxyribonucleic acid preparations into a series of regularly graded fractions of decreasing guanine and cytosine, and rising adenine and thymine contents, but all showing the pairing principles that I have mentioned before. The method used by us was the fractional dissociation, under special conditions, of nucleohistones or histone nucleates. If a preparation

TABLE 27

DIFFERENTIAL ANALYSIS OF PYRIMIDINE DISTRIBUTION IN THREE ANALYTICALLY INDISTINGUISHABLE DNA PREPARATIONS;
PROPORTIONS OF "SOLITARY" PYRIMIDINE NUCLEOTIDES

	Total composition (mole %)			Thymine to cytosine, molar ratio of total	"Solitary units"*			
	Adenine	Guanine	Cytosine		Thymine	(mole % of total constituent)	pTp/pCp molar ratio	
Ox DNA, fraction at 0.75 M NaCl	31.1	19.0	20.0	29.9	1.50	15.9	9.2	2.58
Human DNA, fraction at 1.0 M NaCl	30.9	19.4	19.8	29.9	1.51	19.9	6.3	4.76
<i>Arbacia lixula</i> DNA, fraction at 2.6 M NaCl	31.0	18.5	19.6	30.9	1.58	16.0	14.0	1.80

* pTp and pCp designate the 3',5'-diphosphates of thymidine and deoxycytidine, respectively, which are released by the degradative procedure (0.1 M H₂SO₄, 30 min, 100°).

of deoxyribonucleic acid from mammalian tissue really comprises between 100,000 and 300,000 different individuals, the isolation of "pure" compounds is unattainable; and it is even difficult to conceive of the criteria that would apply. For my part, I must say that the significance of the existence of many deoxyribonucleic acid fractions is by no means clear. There is still, I believe, a great mystery hidden in the structure of the nucleic acids.

From much of what I have said before it must have been evident how important it would be to gain an insight into the order of alignment of the nucleotide monomers in a nucleic acid chain; for the "information" coded into the nucleic acids must find its chemical expression in the specific arrangement of the purines and pyrimidines on the sugar phosphate backbone, that is, in the nucleotide sequence. I have no time here to discuss our earlier attempts, such as the stepwise enzymic degradation or the study of the formation and the breakdown of the apurinic acids²³, which had served to emphasize the arrhythmic character of the polynucleotide chain. But I should like to say a few words about the results of our recent work which is based on the method of differential distribution analysis developed with Shapiro²⁸. This procedure, which consists in the analysis of the breakdown products formed by the controlled, stepwise acid degradation of deoxyribonucleic acids, makes use of the great lability towards acid of the purine glycosides. If we consider a segment of a polynucleotide chain, we may assume that it will contain "solitary" pyrimidine nucleotides flanked by purine nucleotides and *vice versa* as well as "bunched" stretches of varying length consisting exclusively of purine or of pyrimidine nucleotides (compare Fig. 8, p. 90). The first stage of acid degradation, to which we shall limit ourselves today, results in the removal of the purines followed by chain rupture in such a manner as to liberate the "solitary" pyrimidine nucleotides in the form of the nucleoside 3',5'-diphosphates. These, as well as other more complex breakdown products, can be determined quantitatively, furnishing a characteristic feature of the particular deoxyribonu-

cleic acid chain which goes beyond what mere constituent analysis can do. I show in Table 27 a small selection of results obtained with three preparations that are indistinguishable analytically.

It would take me too long to discuss all the implications of this work, but I may state the principal findings. (1) Analytically indistinguishable nucleic acid specimens from different cellular sources exhibit entirely different sequence characteristics. (2) In no case is the nucleotide sequence that to be expected from a random arrangement of monomers. (3) At least 70% of the pyrimidines, and therefore 70% of the purines, occur in tracts of at least three of one kind, *i.e.*, three or more purines followed by three or more pyrimidines, etc. This has a curious consequence in that it restricts the number of possible permutations, though this figure is still enormous. For example, a calf thymus deoxyribonucleic acid consisting of 10,000 nucleotides (molecular weight $3 \cdot 10^6$) could exist in more than 10^{8000} isomers. With the, admittedly oversimplifying, assumption that it consists of trinucleotide tracts of one kind, *i.e.*, adenine–adenine–adenine, etc., the figure for permutational isomers drops from 10^{8000} to 10^{2000} . But you will agree even so that the nucleic acids may, in their sequences, contain enough code-scripts to supply a universe with information.

But do we have enough intelligence to read the information thus offered to us? Not yet, at any rate. There are people who suffer from giddiness when watching the Milky Way for a long time on a clear night. Such people should not look into the cell nucleus. Where else is there so much in so little? And beyond the range of our perception, this millimicrocosm does not end, but there begin galaxies of invisibility. Still, I say to those who come after us: *Scibitis*, you shall know! But only then will the real search begin. In a short time, in a very short time, our science has built a ladder into the heavens; but the first two or three rungs are missing. Let us hope that they will be found in the next hundred years.

REFERENCES

- 1 E. CHARGAFF, *Harvey Lectures, 1956-57*, 52 (1958) 57. [See Chapter 6 of this book.]
- 2 R. J. WILLIAMS, *Biochemical Individuality*, Wiley and Sons, New York, 1956.
- 3 L. LOEB, *The Biological Basis of Individuality*, C. C. Thomas, Baltimore, Md., 1945.
- 4 P. B. MEDAWAR, *Harvey Lectures, 1956-1957*, 52 (1958) 144.
- 5 E. CHARGAFF, *Experientia*, 6 (1950) 201. [See Chapter 1 of this book.]
- 6 E. CHARGAFF, *Advances in Protein Chem.*, 1 (1944) 1.
- 7 A. ROSENBERG AND E. CHARGAFF, *Biochim. Biophys. Acta*, 21 (1956) 588.
- 8 A. ROSENBERG AND E. CHARGAFF, *J. Biol. Chem.*, 232 (1958) 1031.
- 9 H. QUASTLER (Ed.), *Information Theory in Biology*, University of Illinois Press, Urbana, Ill., 1953.
- 10 E. CHARGAFF, *Rend. ist. lombardo sci.*, 89 (1955) 101. [See Chapter 4 of this book.]
- 11 E. CHARGAFF, in *The Origin of Life on the Earth*, U.S.S.R. Academy of Sciences, Moscow, 1957, p. 188. [See Chapter 7 of this book.]
- 12 E. CHARGAFF, in S. GRAFF (Ed.), *Essays in Biochemistry*, Wiley and Sons, New York, 1956, p. 72. [See Chapter 5 of this book.]
- 13 H. T. SHIGEURA AND E. CHARGAFF, *Biochim. Biophys. Acta*, 24 (1957) 450.
- 14 H. T. SHIGEURA AND E. CHARGAFF, *J. Biol. Chem.*, 233 (1958) 197.
- 15 E. CHARGAFF, D. ELSON AND H. T. SHIGEURA, *Nature*, 178 (1956) 682.
- 16 E. VISCHER AND E. CHARGAFF, *J. Biol. Chem.*, 176 (1948) 715.
- 17 E. CHARGAFF, E. VISCHER, R. DONIGER, C. GREEN AND F. MISANI, *J. Biol. Chem.*, 177 (1949) 405.
- 18 D. ELSON AND E. CHARGAFF, *Biochim. Biophys. Acta*, 17 (1955) 367.
- 19 R. LIPSHITZ AND E. CHARGAFF, *Biochim. Biophys. Acta*, 19 (1956) 256.
- 20 A. LOMBARD AND E. CHARGAFF, *Biochim. Biophys. Acta*, 20 (1956) 585.
- 21 A. LOMBARD AND E. CHARGAFF, *Biochim. Biophys. Acta*, 25 (1957) 549.
- 22 T. TSUMITA AND E. CHARGAFF, *Biochim. Biophys. Acta*, 29 (1958) 568.
- 23 E. CHARGAFF, in E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vol. I, Academic Press, New York, 1955, p. 307.
- 24 O. T. AVERY, C. M. MACLEOD AND M. MCCARTY, *J. Exptl. Med.*, 79 (1944) 137.
- 25 E. CHARGAFF, *Symposia Soc. Exptl. Biol.*, 9 (1955) 32. [See Chapter 3 of this book.]
- 26 R. VENDRELY, in E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vol. II, Academic Press, New York, 1955, p. 155.
- 27 E. CHARGAFF, C. F. CRAMPTON AND R. LIPSHITZ, *Nature*, 172 (1953) 289.
- 28 H. S. SHAPIRO AND E. CHARGAFF, *Biochim. Biophys. Acta*, 23 (1957) 451; 26 (1957) 596, 608.

CHAPTER 9

*The Problem of Nucleotide Sequence in Deoxyribonucleic Acids**

1. INTRODUCTION

Now that some of the most tremendous problems in biochemistry have, it is said, been solved successfully—such as the nature of the genetic material, its structure and mode of replication, and even its biosynthesis—our time, tired of so much exertion, may take a deep breath and then, perhaps, decide that it all has to be done over again. I should not dare contradict, for I have always been more impressed by the enormity of the gap between claim and achievement than by the magnitude of the former.

In any event, we have set ourselves a much more modest, but by no means easy, task, namely, the elucidation of some of the aspects of the primary structure of deoxyribonucleic acids, especially in regard to the arrangement of the nucleotide constituents. In this connection, and since this talk is given in Stockholm, I may be permitted to recall that it was here, and also in Uppsala, that I had the first opportunity to review our original observations on deoxyribonucleic acid. This is what I said in 1949 and what was printed a few months later¹.

“...The deoxypentose nucleic acids extracted from different species thus appear to be different substances or mixtures of closely related substances of a composition constant for different organs of the same species and characteristic of the species.

* This chapter has been revised and expanded from a previous publication reprinted with permission from T. W. GOODWIN AND O. LINDBERG (Eds.), *Biological Structure and Function*, Vol. I, Academic Press, London and New York, 1961, p. 67.

"The results serve to disprove the tetranucleotide hypothesis. It is, however, noteworthy—whether this is more than accidental, cannot yet be said—that in all deoxypentose nucleic acids examined thus far the molar ratios of total purines to total pyrimidines, and also of adenine to thymine and of guanine to cytosine, were not far from 1."

You will recognize here, among other things, the first statement of the well-known pairing principles.

Even the earliest observations on the nucleic acids, which showed the existence of remarkable similarities and, at the same time, of outstanding differences in the distribution and, therefore, the sequence of the constituent monomers, the nucleotides, made it appear of great interest to learn something about the structural principles that came into play. I shall first discuss some of the basic concepts that must be considered.

2. REMARKS ON THE CONCEPTUAL BASIS OF SEQUENCE ANALYSIS

a. The problem of homogeneity

What determines the composition of a given biological polymer has been much discussed in the recent past, at any rate in regard to proteins and nucleic acids. Almost no attention has been paid, as has been pointed out recently², to the presumably equally rigid control of the composition of other cell-specific polymers, such as the polysaccharides or even certain macromolecular lipids. The usual expedient, in which I wish I could concur with greater enthusiasm³, has been the formulation of some sort of template which obviously makes up in versatility for what it lacks in concreteness. This Disneyland of templates and pools and feedbacks, presided over by never-smiling augurs, will make us the laughing stock of later times.

What determines the size of a given biological polymer has, on the other hand, been neither determined nor even much discussed. The cell will in many instances contain, or at least it will be made to yield, compounds belonging to the same class, but very much different in size. Proteins, the best investigated group of cellular macromolecules, vary in size over a considerable range,

from something like 12,000 to many millions⁴. Deoxyribonucleic acid preparations appear, in contrast, to fall into two principal groups of molecular weight: (a) of 6 to 8 million; (b) of 12 to 16 million⁵.^{*} Such distinctions are probably not of great heuristic value, since the cell, which is not simply a bag containing many chemicals of different size and quality, must impress on its components a pattern of multiple associations which it is not possible even to define at present.

Dissimilarity in size is, however, less of a problem for the student of the chemistry of deoxyribonucleic acids than is the possible heterogeneity of his specimens as regards their composition and sequence characteristics. It is, in fact, possible to divide a preparation of total deoxyribonucleic acid into a series of fractions of graded composition (starting with a, sometimes extreme, GC type⁷ and going to a very marked AT type) which, however, all still exhibit the pairing regularities⁸; but the real meaning of such a fractionation is far from clear. It could be, as was pointed out by us⁹, that these fractionated preparations actually represent some form of sub-units of a large aggregate, though it is not easy to describe the nature of the links that are disrupted *repeatedly* during the mild fractionation procedure. In this view, in its extreme form, the nucleus of a given species could conceivably contain only one type of deoxyribonucleic acid. At the other extreme would have to be placed the view that a given nucleic acid preparation may comprise an entire spectrum of differently constituted individuals so that no two nucleic acid molecules within the same nucleus would be entirely identical¹⁰. That neither of these opposite extremes can as yet be disavowed entirely demonstrates the deep cleft that still exists between metabiological assertions and scientific facts.

^{*} An interesting exception—a polydeoxyribonucleotide of mol. wt. 10,000 to 12,000—has been described recently⁶ as a component of crystalline cytochrome *b*₂. It is noteworthy that this small polymer of only about 40 nucleotides still exhibits the pairing principles¹ mentioned above.

b. The problem of macromolecular structure

It would, of course, be good if the investigator of the structure of a deoxyribonucleic acid knew whether he was dealing with one chain or with two¹¹ or even more¹² complementary chains; but at the present, little advanced, state of our knowledge of the nucleic acids this is not essential. The value of any information on the sequential characteristics of the nucleic acid is not diminished if the complementary chains are considered to be joined at one end or even at both ends, as may well be the case, so as to constitute an uninterrupted sequential progression. I am not aware of the demonstration of distinct end groups in undegraded deoxyribonucleic acid.

The maintenance of sequence integrity of the preparations to be examined is, on the other hand, a problem of great importance. If—quite apart from the secondary valence forces supporting the architecture of the secondary or tertiary structures of the deoxyribonucleic acids—there really exist weak links in their primary structure¹³, it may be essential to avoid the rupture of the latter, since otherwise valuable features of sequential arrangement may be lost. This applies, of course, even more to the avoidance of chemical or enzymic degradation during the isolation.

It is quite likely that very few of the deoxyribonucleic acid preparations described so far, and none of the models proposed to describe their structure, are representative of the native state. The actual operative entity—though not necessarily amenable to biological testing *in vitro*—possibly is an aggregate of very long polynucleotide chains linked to each other, perhaps by oligopeptide bridges, and bonded to proteins in a spatially unique configuration. If this is so, the problem of heterogeneity, mentioned in the preceding section, is ostensible rather than actual, having been introduced as a necessary, but strictly non-biological, artifact of purification. By this token, a preferred isolation method for deoxyribonucleic acid would be one that avoided, as far as possible, denaturation by chemical or physical means. Nearest to these requirements is perhaps the procedure described

on p. 326 of a previous survey⁷ and applied by us to preparations from calf thymus¹⁴ and *Escherichia coli* protoplasts¹⁵.

c. The problem of statistical sequence analysis

The possibility that no two nucleic acid molecules within the same nucleus are entirely identical offers "a prospect that would seem to condemn us to forced statistics for life"¹⁶. It is this more than anything else that has made work on the nucleotide sequence in nucleic acids appear so unattractive. Who would, after all, undertake to read a book that has been passed through a grinder? Nevertheless, being rather modest in what I expected to gain from a perusal of the nucleic acid text, I have never been able to share these apprehensions completely. I knew that a great deal can be learned about an unknown language through a study of its phonemes, their frequency, distribution density, and allophonic relationships.

If the total deoxyribonucleic acid of a given species represents a text, it is made up of "words"—the individual molecules—that are composed of a singularly meagre alphabet: four or five letters. But the words so spelled out are 10,000-letter words, each of which could occur in a fantastically great number of positional isomers: between 10^{2000} and 10^{8000} , according to how many restrictions on neighbors are admitted³. The situation facing us in examining a nucleic acid preparation comprising a large number of isomers or homologues would, then, be comparable to one in which all the words in a dictionary are lined up end to end in a continuous, and essentially arrhythmic and aperiodic, sequence.

It is quite clear that the first attempt at unraveling such a clutter will have to be based on statistics and that it must limit itself to the description of tendencies or trends of arrangement. To give an example: running together the thirteen words making up the first sentence of *King Lear* I obtain a monster word of fifty-seven letters of which twenty-one are vowels. On this word a number of determinations can be made: (a) the ratio of consonants to vowels; (b) the nature of the individual consonants and vowels; (c) the relative frequency of each constituent. If I

have a way of removing the vowels without disturbing the rest of the arrangement, I shall isolate six solitary consonants, eight pairs of consonants, three bunches of triple consonants, and one cluster of five consonants in a row. Each of these units, I would conclude, was originally flanked on both sides by vowels. Other words would yield other combinations, with the unambiguity of distinction increasing with the length of the consonant clusters. In very long words composed of only two vowels and two or three consonants, unique clusters can be expected only very rarely; but the relative frequency of the various combinations of consonants (runs of 1, 2, 3, etc.) will be a means of unique differentiation, even though it will not yet make it possible to reconstruct the entire text.

This is what we have been doing with many deoxyribonucleic acid preparations.

3. REMARKS ON NOMENCLATURE

a. Definition of the term "homotope"

If one monomeric constituent can take the position of another in a definable segment of a polymer, we propose to designate it as a *homotope* (from the Greek for occupying the same place). If in the A-chain of insulin positions 8 to 10 are occupied by Ala.Ser.Val in the ox¹⁷, but by Ala.Gly.Val in the sheep¹⁸, I would say that serine and glycine are homotopic with respect to this sequence. The importance of this term for a consideration of nucleic acid structure will become clear presently.

b. Definition of the term "pleromer"

If in the total composition of polymers that are characterized by balances such as the well-known pairing principles in deoxyribonucleic acid^{1,7}, and in a more limited way in ribonucleic acids¹⁹, one constituent can ostensibly replace another in respect to these balances, we propose to designate it as a *pleromer* (from the Greek for filling up the measure). Thus, if in the deoxyribonucleic acid of wheat germ²⁰ the molar quantity of guanine

equals that of the sum of cytosine and 5-methylcytosine, I would say that these two 6-amino pyrimidines are pleromeric.

c. Solitary and bunched nucleotides

A pyrimidine nucleotide that, within the polynucleotide chain of a nucleic acid, is contiguous only to purine nucleotides, will be referred to as "solitary"; pyrimidine nucleotides that occur in

Fig. 11. Schematic representation of a fragment of a double strand of a deoxyribonucleic acid. The purines (A, adenine; G, guanine) are depicted in black, the pyrimidines (C, cytosine; T, thymine) in white.

tracts of two or more, flanked on both sides by purine nucleotides, will be designated as "bunched"²¹. In the diagram shown in Fig. 11, solitary thymidylic acid, for instance, appears in positions 3, 9 and 8', bunched pyrimidine nucleotides are seen in positions 5-7.

4. EARLY ATTEMPTS AT SEQUENCE INVESTIGATION

Since 1947, when in an early review on deoxyribonucleic acids (p. 32 of Ref. 22), I first discussed the possible importance of variations in the nucleotide sequence for the biological specificity of a deoxyribonucleic acid, this problem has remained among the

interests of my laboratory. Our first chemical evidence on the existence of different deoxyribonucleic acids^{23, 24} made it clear that whatever "code" was carried by the nucleic acids must be imprinted on the sequential arrangement of the monomeric constituents and prompted us to search for ways to approach this difficult task.

The general problem is similar to that of the structure of proteins, but immensely more forbidding. The very great length of the chains composed of a much smaller number of different monomers raises obstacles that render improbable an unambiguous solution by existing methods. Moreover, the available procedures are not yet as refined as in the case of the proteins: few specific enzymes, no generally applicable method for marking the end groups. On the other hand, the remarkable difference in the stability of the glycosidic bonds holding the purines and the pyrimidines offered a novel possibility that we have exploited fully.

Our first attempts in the direction of utilizing a specific enzyme were concerned with the problem of a recognizable repeating unit.

a. Stepwise degradation by deoxyribonuclease

When crystalline pancreatic deoxyribonuclease acts on a highly polymerized preparation of a deoxyribonucleic acid, it is possible to conduct the experiment in such a manner as to collect the products detached gradually in a dialyzable form and to separate them from the enzyme-resistant core^{7, 25}. The latter is characteristically different in composition from the intact polymer; there is, moreover, a significant trend in the composition of the products produced by the stepwise enzymic digestion. No indications of any regularity in the release or the composition of fragments were observed; no sub-units of recognizably recurrent structure were seen: deoxyribonucleic acids apparently exhibit a largely arrhythmic nucleotide sequence.

b. Apurinic acid

This has proved a very useful compound; it is, in fact, the first

intermediate product in all procedures involving acid degradation²⁶. When a deoxyribonucleic acid is exposed to mildly acidic conditions (pH 1.6 at 37°) all the purines are cleaved off gradually, leaving behind a polymer representing the original polynucleotide, but composed of deoxyribophosphate units, at the places of the previous purine nucleotides, and of the pyrimidine nucleotides in unchanged ratio and at the same position as in the native starting material. For instance, in the segment represented in Fig. 11, positions 3, 5-7, 9, 1', 2', 4', 8', 10' would remain unchanged, whereas in the remaining places the purines would now have made room for the reactive aldehydo groups of the deoxy sugar. The presence of the aldehydo group²⁷ and, therefore, of a free hydroxyl at 4' brings about a remarkable labilization of the polymer: it is broken not only by alkali, but even by buffers (pH 8.6) containing primary amino groups²⁸. It is likely that this susceptibility of the poly-sugar-phosphate backbone of apurinic acid to amines is involved in the degradation reactions employing diphenylamine under acidic conditions²⁹.

Studies on the arrangement of the pyrimidine nucleotides in apurinic acid²⁸ were, in fact, the first that made possible an approach to the problem of nucleotide sequence in deoxyribonucleic acid. The information so obtained was limited, since it was only qualitative, but it showed that a considerable portion of the pyrimidine nucleotides, and therefore also of the purine nucleotides, was arranged in the form of tracts of several nucleotides of one kind.

5. SEQUENCE STUDIES THROUGH DIFFERENTIAL DISTRIBUTION ANALYSIS

a. Mechanism

The important discovery that among the fragments produced by the acid degradation of deoxyribonucleic acids there are found the 3',5'-diphosphates of deoxycytidine and thymidine is due to the work of Levene, Thannhauser and their collaborators³⁰⁻³³. A more recent re-investigation³⁴ by improved techniques, which

confirmed the occurrence of these diphosphates, prompted a discussion of the possible bearing of this finding on the problem of nucleic acid structure (p. 366 of Ref. 7). It was clear that for these fragments to serve as a tool in research on nucleotide sequence the mechanism of their release had to be better understood. It was also indispensable to be able to study them under rigorously controlled conditions and on a quantitative basis.

The kinetics of the liberation of the pyrimidine nucleoside diphosphates were first studied, in collaboration with Shapiro, on a series of models, namely, the several deoxyribo dinucleotides all having cytidylic acid as one of their components³⁵. These investigations served as the basis of a well-controlled, differential hydrolysis procedure which was usually performed in three stages^{21, 36}. The diagnostically most valuable results are obtained in stage I (0.1 *M* H₂SO₄, 35 min, 100°), when the release of pyrimidine nucleoside diphosphates can be taken to reflect directly the abundance of these units as solitary pyrimidine nucleotides in the polynucleotide chain. In the present survey I shall limit myself to the findings based on this stage of the differential hydrolysis procedure. The results yielded in stages II and III, which involve longer periods of hydrolysis, are mainly indicative of the secondary cleavage of bunched pyrimidine sequences; they are of great value as further means of differentiation between analytically indistinguishable nucleic acids of different cellular origin. This will be exemplified in the following section.

In its simplest form the cleavage of a polynucleotide chain and the formation of the diphosphates of the pyrimidine nucleosides (Py) may be regarded as a series of β elimination reactions²¹. After the liberation of the purines the first elimination pre-

sumably occurs at the broken line A. Since no extraneous sugar fragment is found in ester linkage with the resulting nucleoside diphosphate, a subsequent hydrolysis or elimination must take place at the other flanking sugar (broken line C). It is likely that this cleavage follows a β aldehyde elimination (broken line B), which has left a double bond between the second and third carbon atom of the sugar. Similar considerations apply to the

TABLE 28

DIFFERENTIAL DISTRIBUTION ANALYSIS OF THREE OTHERWISE
INDISTINGUISHABLE DEOXYRIBONUCLEIC ACID PREPARATIONS*

Source:	Ox			Man			Arbacia lixula		
DNA fraction:	(0.75 M)			(1.0 M)			(2.6 M)		
	T	C	TC	T	C	TC	T	C	TC
Total pyrimidine, as mole % P	29.9	20.0		29.9	19.8		30.9	19.6	
Solitary pyrimidine, as mole % of total constitu- ent in DNA	15.9	9.2		19.9	6.3		16.0	14.0	
Solitary or bunched pyrimidines, as mole % P	4.75	1.84	1.72	5.95	1.25	1.71	4.94	2.74	1.66
Total T/C, molar ratio		1.50			1.51			1.58	
Solitary T/C, molar ratio		2.58			4.76			1.80	

* The figures are taken from previous papers^{21, 36}. The deoxyribonucleic acid fractions are described by the molarity of NaCl employed for the dissociation of the histone nucleate⁹. T designates thymine or thymidylic acid, C cytosine or deoxycytidylic acid, TC the two isomeric dinucleoside triphosphates comprising cytidine and thymidine (reported as moles of dinucleoside triphosphate per 100 gram-atoms of nucleic acid phosphorus).

release of bunched pyrimidine sequences. A more than tentative formulation of the reaction mechanism will, however, have to await a better understanding of the fate of the free deoxyribose residues liberated in the initial stage of the removal of purines by acid.

b. Nucleotide arrangement in analytically indistinguishable deoxyribonucleic acids of different cellular origin

It happens quite often that deoxyribonucleic acids of taxonomically entirely different origin exhibit identity of composition as regards the distribution of the nitrogenous constituents (compare the example of the deoxyribonucleic acids of the sheep and the salmon cited previously¹⁶). In such cases, the method of differential distribution analysis discussed in this survey is destined to be of great value. It permits the examination of detailed features of the nucleotide arrangement, far beyond the possibilities of distinction offered by total constituent analysis. I have selected, for inclusion in Table 28, three examples of nucleic acid fractions that cannot be distinguished analytically. There can be little doubt about the entirely different sequential plans governing the structure of these polymers if the values for the abundance of solitary pyrimidine units are compared.

c. Nucleotide arrangement in different deoxyribonucleic acids of animal origin

Table 29 compares the distribution findings obtained with total, unfractionated preparations of the deoxyribonucleic acids from calf thymus, human spleen, and sperm preparations from two sea urchin species. These data do not exhaust the information gathered by us on different animal nucleic acids²¹, but are presented here to indicate the scope of the method.

d. Nucleotide arrangement in the total deoxyribonucleic acid of rye germ and in a complete series of fractions

The use of a deoxyribonucleic acid from a plant source offers several advantages. These nucleic acids have been shown to con-

TABLE 29
DIFFERENTIAL DISTRIBUTION ANALYSIS OF ANIMAL DEOXYRIBONUCLEIC ACIDS*

Source:	Ox			Man			Arbacia lixula			Paracentrotus lividus		
Organ:	Thymus			Spleen			Sperm			Sperm		
	T	C	TC	T	C	TC	T	C	TC	T	C	TC
Total pyrimidine, as mole % P	28.1	20.6	28.9	20.4	30.5	19.2	32.1	17.3				
Solitary pyrimidine, as mole % of total constituent in DNA	15.4	9.7	13.2	8.2	17.6	14.1	16.1	15.5				
Solitary or bunched pyrimi- dines, as mole % P	4.33	2.00	1.44	3.81	1.67	1.35	5.37	2.71	1.43	5.17	2.68	1.52
Total T/C, molar ratio		1.36		1.42		1.59		1.86				
Solitary T/C, molar ratio		2.17		2.28		1.98		1.93				

* All nucleic acids analyzed were total, unfractionated preparations. The figures are taken from a previous paper²¹. For other explanations, compare Table 28.

tain a fifth nitrogenous constituent, 5-methylcytosine, in appreciable quantity^{7, 20}. In the specimens isolated from both wheat and rye germ the concentration of this pyrimidine amounts to 5.9 mole % (Table 30). This means that nearly a quarter of the cytosine has been "replaced" by methylcytosine. Since the molar concentration of guanine equals the sum of these two pyrimidines, one may consider them as pleromeric in the sense defined above. Minor constituents of deoxyribonucleic acid (or, for that matter, of ribonucleic acid) are destined to play an important role in future work on the elucidation of nucleic acid structure, as they will serve as additional markers in the array of nucleotides constituting the polymer chain. Furthermore, the presence of a pyrimidine sharing with cytosine the property of being a 6-amino derivative and with thymine that of being a 5-methylpyrimidine is of the greatest interest for a better understanding of the mechanisms of selection, incorporation, and replication that are at work.

TABLE 30

COMPOSITION OF DEOXYRIBONUCLEIC ACIDS OF RYE GERM AND WHEAT GERM

	<i>Rye germ*</i> <i>Moles/100 gram-atoms P</i>	<i>Wheat germ**</i> <i>Moles/100 gram-atoms P</i>
Adenine	27.7	28.1
Guanine	22.6	21.8
Cytosine	16.4	16.8
Methylcytosine	5.9	5.9
Thymine	27.4	27.4
<i>Molar ratios</i>		
Adenine + thymine to guanine + cytosine + methylcytosine	1.23	1.25
Purines to pyrimidines	1.01	1.0
Adenine to thymine	1.01	1.03
Guanine to cytosine + methylcytosine	1.01	0.96
Cytosine to methylcytosine	2.78	2.85

* Taken from a previous publication³⁷.** Taken from a previous publication³⁸.

Before the study of the nucleotide arrangement in the deoxyribonucleic acid of rye germ could be undertaken, it was necessary to examine the properties of deoxymethylcytidine 3',5'-diphosphate whose production is indicative of the frequency of this pyrimidine as a solitary nucleotide³⁹. At the same time, the two dinucleoside triphosphates comprising 5-methylcytidine and either cytidine or thymidine were also studied.

The investigation of the deoxyribonucleic acid of rye germ³⁷ offered us, for the first time, the opportunity of comparing the pyrimidine distribution in the total preparation with that found in a complete series of seven fractions isolated through the fractional dissociation of the histone nucleate. These fractions were, in their distribution and composition, remarkably similar to those previously isolated by the fractionation of the deoxyribonucleic acid from wheat germ³⁸. The availability of such a complete series of fractions permitted, moreover, a test of the validity of the results given by the differential distribution analysis; a weighted average of the values recorded for the individual fractions could be compared with those given by the total nucleic acid preparation that had not been subjected to fractionation and showed an excellent agreement³⁷.

The principal results of this study are, as regards the frequency of solitary pyrimidine nucleotides, shown schematically in Fig. 12. In this histogram, we give a brief description of the total pyrimidine composition of all specimens examined—total deoxyribonucleic acid and seven fractions—and depict the quantities of solitary nucleotide of each of the three pyrimidines as per cent of the total constituent in the preparation. 5-Methylcytosine, it will be seen, has a surprisingly high tendency, when compared with the other 6-aminopyrimidine, cytosine, to occur as a solitary unit, similar in this respect to thymine. I shall return to this point later.

e. Nucleotide arrangement in a microbial deoxyribonucleic acid of the GC type

Our survey has so far all been dealing with deoxyribonucleic

acids that belong to the AT type. It was of interest to see whether the same considerations could be applied to a nucleic acid of the GC type, *i.e.*, one in which guanine and cytosine preponderated. We first discovered this type in tubercle bacilli²⁴; but it has since that time been found in many micro-organisms^{7, 41}. The distribution analysis of the deoxyribonucleic acid of an avirulent variant (BCG) of bovine tubercle bacilli is shown in Table 31. In the nucleic acids of the AT type, as was shown before, the quantity of thymidine 3',5'-diphosphate released

Fig. 12. Pyrimidine distribution analysis of the total deoxyribonucleic acid of rye germ and its fractions. The first column indicates the NaCl molarity at which dissociation of the fraction from its histone salt occurred. The second column summarizes the composition of the specimens in terms of the molar concentrations (per 100 gram-atoms of nucleic acid P) of thymine (T), cytosine (C), and 5-methylcytosine (M). The histogram shows the frequency of the solitary nucleotides, isolated as the nucleoside 3',5'-diphosphates, relative to the total concentration of the constituent in the nucleic acid. Based on previously published data³⁷.

during hydrolysis exceeded that of deoxycytidine 3',5'-diphosphate: this indicated the occurrence of larger quantities of solitary thymine than cytosine, but the ratios of the solitary residues were not those that would have been predicted statistically from the total abundance of these pyrimidines in the polymers. In the nucleic acid of the GC type an inversion in the rates of liberation of the nucleoside diphosphates is observed, as expected; but in this instance, the ratios of solitary and of total pyrimidines are not widely different: 0.59 *vs.* 0.52.

TABLE 31

DIFFERENTIAL DISTRIBUTION ANALYSIS OF DEOXYRIBONUCLEIC ACID OF BCG*

	<i>T</i>	<i>C</i>	<i>CC</i>
Total pyrimidine, as mole % P	16.7	32.3	
Solitary pyrimidine, as mole % of total constituent in DNA	13.6	12.0	
Solitary or bunched pyrimidines, as mole % P	2.27	3.88	1.13
Total T/C, molar ratio		0.52	
Solitary T/C, molar ratio		0.59	

* The figures are taken from a previous paper⁴⁰. CC designates dicytidine triphosphate. Compare Table 28 for other explanations.

f. Nucleotide arrangement in the deoxyribonucleic acid of E. coli grown in the absence and presence of 5-bromouracil

The deoxyribonucleic acid of *E. coli* is unusual, being composed of nearly equimolar proportions of the four nucleotide constituents^{7, 41}. No obvious distinctions appear, even when different variants and mutants are examined, although it is not unlikely that a careful differential distribution analysis would reveal some differences. Such a study has, however, not yet been undertaken.

It appeared of great interest to study the effect of a fraudulent analogue on the sequence characteristics of a deoxyribonucleic

acid. For this reason, the nucleic acid produced by thymine auxotrophs of *E. coli*, grown in the presence of thymine, was investigated and compared with that formed by the same organisms when supplied also with 5-bromouracil. This 6-ketopyrimidine is known to be accepted by certain bacterial cells as an analogue of thymine⁴²⁻⁴⁴. The incorporation of the halopyrimidine, which is essentially limited to cells unable to synthesize thymine, is accompanied by many biological consequences. The cells suffer a gradual loss of viability⁴⁵ and exhibit a considerable increase of the rate of formation of mutants⁴⁶, sometimes with preferred sites of particularly high mutability⁴⁷. It has also been shown that bacteriophage T₂ particles may retain their infectivity, despite the uptake of a high proportion of 5-bromouracil⁴⁸.

The results of our first study with the thymine auxotroph, strain I, were quite surprising⁴⁴: they seemed to indicate a very severe distortion of the sequence characteristics of the deoxyribonucleic acid in which about 35 mole % of thymine had been replaced by 5-bromouracil⁴⁹. Further work with this and another mutant strain has, however, failed to bear out this conclusion;

Fig. 13. Chromatography of hydrolysates of deoxyribonucleic acid of *E. coli*, mutant strain I (Preps. 3 and 4 in Table 32): (a) grown in the absence; (b) in the presence of 5-bromouracil. Elution from DEAE-cellulose with an increasing linear gradient of LiCl in lithium acetate buffer (pH 5); other techniques as in Ref. 51. Total volume of eluent 1,350 ml; 7-ml fractions. Taken from a recent paper⁵⁰.

TABLE 32

COMPOSITION OF DEOXYRIBONUCLEIC ACID OF TWO THYMINE AUXOTROPHS OF *E. coli* GROWN IN THE ABSENCE AND PRESENCE OF 5-BROMOURACIL*

Mutant strain	Preparation No.	Growth supplements	Moles constituent per 100 moles total				Molar ratios			
			A	G	C	T	B	(A + T + B)/(G + C)	6-Am/6-K	A/T or A/(T + B) B:(T + B)·100
15 t-arg-	1	T	25.0	24.9	25.6	24.4	—	0.98	1.03	1.02
	2	T + B	24.4	25.6	25.7	10.6	13.7	0.95	1.00	1.00
I	3	T	24.6	25.5	25.6	24.3	—	0.96	1.01	1.01
	4	T + B	24.9	25.1	25.4	13.5	11.1	0.98	1.01	1.01
I	5	T	24.8	25.0	25.5	24.8	—	0.98	1.01	1.00
	6	T + B	25.3	24.3	25.1	16.4	9.0	1.03	1.01	1.00

* Based on previously published results^{49,50}. The thymine auxotroph 15 t-arg- was grown on a synthetic medium supplemented, per ml, with 1 μ g of thymine or with 1 μ g of thymine and 10 μ g of 5-bromouracil, for two cycles of replication at 37°; strain I was similarly grown for 18 h at 37° in the presence of 2 μ g/ml of thymine, with or without the addition of 10 μ g/ml of the halopyrimidine. Abbreviations: A, adenine; G, guanine; C, cytosine; T, thymine; B, 5-bromouracil; 6-Am, 6-amino constituents (A, C); 6-K, 6-keto constituents (G, T, B).

and I shall here consider the situation as it appears in the light of our most recent evidence⁵⁰. When the results of the total analysis of six nucleic acid specimens from two mutant strains are compared (Table 32), it will indeed be seen that 5-bromouracil and thymine are pleromeric with respect to the equality of the molar quantities of purines and pyrimidines, of 6-amino and 6-keto nucleotides, and of 6-aminopurine (adenine) and 6-ketopyrimidines (thymine + 5-bromouracil).

The improved methods for the determination of the distribution of pyrimidine clusters⁵¹ were applied to Preparations 1-4 (Table 32). The elution diagram showing the separation of the pyrimidine clusters released from Preparations 3 and 4 is shown in Fig. 13, the phosphorus distribution in all four preparations of nucleic acid in Table 33. It may be seen (Fig. 13) that the nucleic

TABLE 33

DISTRIBUTION OF PYRIMIDINE NUCLEOTIDE CLUSTERS IN DEOXYRIBONUCLEIC ACID OF TWO THYMINE AUXOTROPHS OF *E. coli* GROWN IN THE ABSENCE AND PRESENCE OF 5-BROMOURACIL⁵⁰

Strain:		15 <i>t-arg</i> ⁻		<i>I</i>	
Growth supplements:		<i>T</i>	<i>T + B</i>	<i>T</i>	<i>T + B</i>
DNA Prep. No.:		1	2	3	4
Fraction No.	Number of pyrimidine nucleotides in fragment (<i>n</i>)	Proportion of DNA phosphorus, %			
I	1	13.7	13.4	13.9	13.9
II	2	12.1	12.3	12.2	11.7
III	3	7.7	7.6	7.4	7.6
IV	4	5.4	5.5	5.1	5.6
V	5	3.4	3.6	3.7	4.0
VI	6	2.2	8.1**	2.3	6.5**
VII	7	1.3		1.4	
VIII	8	1.1		1.2	
IX		3.3*		3.5*	
Total		50.2	50.5	50.7	49.3

* Based on an average chain length of $n = 11$.

** Based on an average chain length of $n = 8$.

TABLE 34

FREQUENCY OF SOLITARY AND COUPLED PYRIMIDINE NUCLEOTIDE UNITS IN DEOXYRIBONUCLEIC ACID OF A THYMINE AUXOTROPH OF *E. coli* t-arg- IN THE ABSENCE AND PRESENCE OF 5-BROMOURACIL*

<i>Growth supplements:</i>		<i>T</i>	<i>T + B</i>
<i>DNA Prep. No.:</i>		<i>1</i>	<i>2</i>
		<i>Moles of pyrimidine per 100 gram-atoms of DNA phosphorus</i>	
Solitary units	Tp ₂	3.81 (0.05)	1.60 (0.05)
	Bp ₂	—	2.02 (0.08)
	Cp ₂	4.53 (0.17)	4.84 (0)
	Total T or B + T	3.81	3.62
Coupled units	T ₂ p ₃	2.42 (0.01)	0.66 (0.10)
	TBp ₃	—	1.32 (0.24)
	B ₂ p ₃	—	0.97 (0.05)
	TCp ₃	5.42 (0.06)	2.50 (0.14)
	BCp ₃	—	3.18 (0.10)
	C ₂ p ₃	3.44 (0.02)	3.27 (0.11)
Total T or B + T		5.13	5.79
B/(T + B) · 100 in solitary units		—	56
B/(T + B) · 100 in coupled units		—	56
B/(T + B) · 100 in total DNA		—	56

* The nucleic acid preparations are numbered as in Table 32. The same abbreviations as in Table 32 are used, with p indicating the number of phosphate groups in the fragments having the general formula Py_np_{n+1} (Py = pyrimidine). No distinction is made between sequential isomers of the coupled units. The figures are the averages of the results of two separate hydrolysis experiments, each chromatographic component being determined twice; the deviations from the averages are given in parentheses. This table is taken from a recent publication⁵⁰.

acid specimens produced in the absence of the halopyrimidine yielded at least eight distinct fractions, going from solitary units to octanucleotide sequences, whereas with the preparations containing 5-bromouracil the longest runs separated clearly were those of five pyrimidines in a row. It may be that this often observed difference is not without significance.

In all nucleic acid specimens the frequency of solitary and coupled pyrimidine nucleotide units—and in some cases also that of triplets—was determined. An example of such an analysis is shown in Table 34. It will be seen that—in contrast to the first observations⁴⁹—no evidence emerged of a gross non-randomness in the ability of the fraudulent analogue to substitute for thymine within the polynucleotide chain.

6. SUMMARIZING REMARKS

a. Sequence characteristics common to all deoxyribonucleic acids

Though the first application of the procedure for the differential analysis of pyrimidine distribution served to demonstrate that deoxyribonucleic acids from different species, even those showing identical base composition, were vastly different with regard to the alignment of their nucleotide constituents (Table 28), the more important and interesting uses of this method bear on other aspects of the sequence problem. These relate, in fact, to the existence of sequential patterns that all deoxyribonucleic acids examined heretofore appear to have in common.

The first consequence of our studies may be the statement that in no case can the nucleotide sequence be predicted from the knowledge of the frequency of the individual constituents. The discarded tetranucleotide hypothesis is a primitive example of an attempt to make such a prediction; I have dealt with this a long time ago¹. This is, incidentally, quite similar to what one would observe in a meaningful text: a certain letter may be particularly frequent; but this does not mean that it will occur, with commensurate frequency, as a doublet. We shall probably find, in general, that the symbols constituting a text are not arranged at random, if this term is understood to refer to an alignment governed only by the relative frequency of the elements, once the requirement of large numbers is fulfilled.

This brings up the problem of the non-randomness of the nucleotide sequence in deoxyribonucleic acids which we have discussed fully in a previous paper²¹. Whether strict randomness

is conceivable in a living system capable of orderly replication, at any rate on the level of macromolecular events, is a question which it would be out of place to discuss here. The entire complex of randomness in biology would certainly deserve a more profound examination than I am equipped to perform. In any event, the conclusion that in no case investigated the pattern of nucleotide distribution was that to be expected from a random arrangement of constituents^{21, 35, 37, 40} may contribute to the assurance that in studying the nucleotide sequence in a deoxyribonucleic acid we are, indeed, dealing with some sort of meaningful message. If the array of constituents in a biologically functional polynucleotide has been compared to a giant throw of dice, we must conclude that it is a throw of loaded dice.

Another consequence of our studies on deoxyribonucleic acids of animal and plant origin^{21, 37, 51} is the conclusion that at least 60% of the pyrimidines occur as oligonucleotide tracts containing three or more pyrimidines in a row; and a corresponding statement must, owing to the equality relationship, apply also to the purines.

b. Possibilities of cluster analysis

It is quite clear that the various solitary nucleotides and the small bunched oligonucleotides, such as doublets or triplets, will vary in frequency, *i.e.*, in relative quantity, according to the plan governing the nucleotide sequence of the particular deoxyribonucleic acid; but they are most unlikely to exhibit uniqueness of quality and will, therefore, not be sufficient to define a sequence unambiguously. Uniqueness may, however, be expected, at least occasionally, when very large clusters can be included in the survey. We have made a beginning with regard to clusters of pyrimidine nucleotides⁵¹. When a deoxyribonucleic acid is cleaved by acid under moderate conditions (0.1 M H₂SO₄, 35 min, 100°) and the fragments are fractionated on DEAE-cellulose by means of a lithium chloride gradient, the phosphorylated pyrimidine oligonucleotides can be separated in order of increasing chain length (compare Fig. 13). Further separation within the different

size groups then is based on two-dimensional chromatography on filter paper.

In this manner, essentially the entire pyrimidine phosphorus of a nucleic acid can be accounted for (see the examples in Table 33). With the amounts usually employed for hydrolysis (20 mg of nucleic acid) the isolation of sequences longer than nine is difficult; but there is no doubt that with more material and larger columns the separation of even longer units will be possible, especially if by the use of radioactive labels the sensitivity of the detection of fractions could be extended. What will become, however, increasingly harder or even impossible, is the separation of very large clusters into individual components, *e.g.*, the discrimination between $T_{15}P_{16}$ and $T_{14}Cp_{16}$, etc., let alone the preparation of the many sequential isomers of this type.

One gains the impression that almost all possible combinations of pyrimidine nucleotide units are encountered. But there is one important, and possibly quite significant, exception. There seems to exist a certain bias in the arrangement of homologous pyrimidine clusters: many nucleic acids appear to contain much longer all-thymine than all-cytosine units. This apparent trend may be exemplified by the instance of the deoxyribonucleic acid of *E. coli* which contains nearly equimolar quantities of thymine and cytosine (compare Table 32). One should have expected similar proportions of the stretches of varying length comprising only one or the other of these pyrimidines. Actually, no cytosine units longer than C_3 have been observed in the hydrolysates, whereas T_4 and T_5 units have invariably been found, together with the various mixed fragments such as T_3C , T_2C_2 , TC_3 , T_4C , T_3C_2 and T_2C_3 . In general, cytosine is found in a relatively greater abundance among the smaller units; the converse is true of thymine.

These and related methods, employing specific enzymes, will doubtless have to be refined considerably, before more definitive answers to the sequence problem and its biological meaning become possible. It would, for instance, be of the greatest interest to determine the nature of the lethal sequence of a parasitic

deoxyribonucleic acid that enables it to impose itself upon, and to paralyze, as it were, the bacterial host.

c. What is meant by "replacement"?

The platitude that "nobody is irreplaceable"—if it ever holds true—certainly cannot hold for the deoxyribonucleic acids, if they carry the biological functions ascribed to them currently. Nevertheless, one often finds the statement in the literature that an analogue of a purine or pyrimidine, not normally occurring in nature, is, when offered to the cell, incorporated into a deoxyribo- or ribonucleic acid "in the place" of a particular natural constituent of the nucleic acid. Such statements are, of course, derived from the existence of the pairing principles¹, *i.e.*, they are based on the fact that certain nucleic acid constituents are pleromers, to use the nomenclature suggested above. Good examples of such pleromerism can be seen in Table 30 with respect to cytosine and 5-methylcytosine and in Table 32 with respect to thymine and 5-bromouracil; a more comprehensive selection of such instances will be given at the end of this paper.

What does the assertion that X can be "replaced" in part by Y actually signify? (1) It means that Y is incorporated into the deoxyribonucleic acid chain and that the cell, hence, must possess, or acquire, the enzymic apparatus necessary for the production and utilization of the derivative of Y that is the direct precursor of the polynucleotide chain. (2) If X and Y share the same functional group (*e.g.*, 6-keto or 6-amino) through which alone the selection of the members of the newly forming chain takes place—as foretold by a mechanism postulating the replication of the complementary halves of a double strand⁵²—it follows that every X molecule along the chain has an equal chance of being "replaced" by a Y molecule.

d. Selection of a natural analogue

When we deal with a natural analogue, such as the 5-methyl derivative of cytosine, it is, of course, meaningless to speak of the replacement of one by the other. Such a statement acquires

meaning only in reference to the possible mechanism of the biosynthesis of the polynucleotide chain or to the existence of such principles of pleromerism as those mentioned before. In other respects, in polymer chains that are not constructed haphazardly, each monomeric constituent must have its preordained place; a place that it must have entered, at the time of polymerization, through some form of a specific selection mechanism. A replication mechanism of this nature is, indeed, offered by the scheme⁵² postulating a specific pairing of a 6-aminopurine (adenine) with a 6-ketopyrimidine (thymine) and of a 6-ketopurine (guanine) with a 6-aminopyrimidine (cytosine).

This postulate and the fact that 5-methylcytosine is pleromeric with cytosine could have led to the expectation that these two pyrimidines could replace each other at random in the deoxyribonucleic acid chain, selection having regard only to the presence of a 6-amino group in the pyrimidine. The amazingly constant quantity of methylcytosine incorporated into the deoxyribonucleic acids of wheat and rye germ (Table 30, and compare also the tabulation given in a previous paper³⁸) should, however, render such a conclusion most questionable. In what manner could a guanine residue in the preformed chain exercise a preference for pairing with cytosine, rather than with methylcytosine, in the newly forming counterpart chain? There are surely limits to the amount of foresight that we can attribute to an ostensibly automatic process.

Another observation, made repeatedly^{8, 37, 38}, has to do with the trend of the distribution of 5-methylcytosine in consecutive fractions obtained by the fractionation of deoxyribonucleic acid from calf thymus, wheat germ, and rye germ. In all instances, the relative abundance of 5-methylcytosine with respect to cytosine varied in such a manner that the percentage contribution of the minor pyrimidine was higher in the earlier fractions. In no case could one have claimed that the two 6-aminopyrimidines replaced each other at random within the polynucleotide chain. I have discussed, in more detail, the significance of the finding of a disproportionate distribution of cytosine and 5-methylcytosine in

view of current replication hypotheses at a previous occasion¹⁰.

An even more impressive difference between the ways in which methylcytosine and cytosine are selected for insertion into the polynucleotide emerges from the study of the frequency of these pyrimidines as solitary and bunched nucleotides³⁷. As will be seen in Fig. 12, a remarkably large proportion of methylcytosine, and a surprisingly small proportion of cytosine, exist as solitary units. These two 6-aminopyrimidines can hardly share a common pattern of nucleotide sequence. What is, however, remarkable is that the mole fractions of 5-methylcytosine and thymine appearing as solitary units are nearly equal in most instances. One must conclude that, whereas there is evidence that the two 6-aminopyrimidines are not treated indiscriminately by the selection mechanism, there is some indication that no such distinction is exercised in regard to the two 5-methylpyrimidines. In the terms defined above, cytosine and 5-methylcytosine are pleromers, but not homotopes; thymine and 5-methylcytosine appear to be homotopic with respect to the sequence purine-pyrimidine-purine, although they are not pleromeric.

e. Selection of a fraudulent analogue

It is striking that almost all extraneous purine or pyrimidine analogues that can, by the living cell, be incorporated massively into nucleic acids—*e.g.*, 5-bromo-, 5-chloro- or 5-iodouracil into deoxyribonucleic acid, 5-fluorouracil or 8-azaguanine into ribonucleic acid—belong to the 6-keto type. In the instance investigated in the greatest detail, *viz.*, the incorporation of 5-bromouracil into the deoxyribonucleic acid of *E. coli*, the problem of replacement can be posed directly. In the case of a natural satellite, such as 5-methylcytosine, the complement appears to be fixed rigidly by the cell, whereas no obvious cellular mechanism, except the death of the cell, can be discerned that would limit the extent of incorporation of the halopyrimidine. One would expect the cell to discriminate between a foreign intruder and the normally required thymine; but the auxotroph apparently can be inveigled into accepting the substitute, very much in contrast to the wild type.

The inspection of Table 34 will show that in the nucleotide sequences under comparison the relative contribution of 5-bromouracil to the sum of 6-ketopyrimidines equalled the relative abundance of the analogue in the total nucleic acid. For instance, in Prep. 2 the halopyrimidine accounted for 56% of the sum of the 6-ketopyrimidines of the nucleic acid (Table 32) and the same figure was found for its relative frequency in the solitary and dinucleotide units (Table 34). Similar results were obtained with the other mutant strain, not listed here in detail; and in this case the agreement extended even to the trinucleotides⁵⁰. In addition, the frequencies of thymine in the various pyrimidine nucleotide clusters released from the nucleic acid specimens synthesized in the presence of thymine alone were quite similar to the frequencies of the sum of thymine and 5-bromouracil in the corresponding fractions from preparations produced in the presence of the analogue.

f. The concepts of pleromerism and homotopy and the neighbor problem

The definitions of the suggested terms "pleromer" and "homotope" have been given above. The compilation in Table 35, by no means complete, gives a few examples of such pleromeric relationships in deoxyribonucleic acids. There can be little doubt that in all these instances the established pairing relationships would have been destroyed, had the minor pleromer been omitted from the computations. Only one satellite normally occurring in nature, viz., 5-methylcytosine, could be included in the table. Another minor component—if it can be classified as a normal constituent—namely, *N*-methyladenine, plays a baffling role. It has been reported to occur, among other instances, in the deoxyribonucleic acid of a thymine auxotroph of *E. coli* in not inconsiderable quantity⁵³. Under special conditions of growth, its concentration increases, somewhat in measure with a decrease in the amount of thymine. It is, however, at the moment difficult to see through what mechanism a 6-ketopyrimidine and a 6-methylaminopurine could function as pleromers.

TABLE 35
EXAMPLES OF PLEROMERIC RELATIONSHIPS IN DEOXYRIBONUCLEIC ACIDS

Source of DNA	Pleromers		Molar ratios		Reference
	Analogue or natural satellite (mole %)	Normal pyrimidine (mole %)	Complementariness ratio*	6-Am/6-K**	
Wheat germ	5-Methylcytosine 5.9	Cytosine 16.8	0.96	1.03	38
Rye germ	5-Methylcytosine 5.9	Cytosine 16.4	1.01	1.00	37
<i>E. coli</i>	5-Chlorouracil 14.2	Thymine 7.9	1.06***	1.00***	45
<i>E. coli</i>	5-Bromouracil 9.0	Thymine 16.4	1.00	1.01	49
<i>E. coli</i>	5-Iodouracil 4.1	Thymine 19.6	1.01	0.99	45
Phage T _{2r} ⁺	5-Bromouracil 16.1	Thymine 15.9	1.01	0.97	45

* In the case of 6-aminopyrimidines (cytosine, 5-methylcytosine), this is the molar ratio of guanine to the sum of 6-aminopyrimidines; in the case of the halopyrimidines, the complementariness ratio is the molar ratio of adenine to the sum of 6-ketopyrimidines.

** Molar ratio of 6-amino nucleotides to 6-keto nucleotides.

*** N-Methyladenine (1.4 mole %) omitted from the calculations.

Pleromeric relationships in ribonucleic acids are much more difficult to establish, as the only complementariness* useful for scrutiny is the equality of the 6-amino and 6-keto nucleotides¹⁹. Of the three examples listed in Table 36 only the one dealing with the incorporation of 5-fluorouracil⁵⁶ carries conviction. It is not impossible that, when more careful analyses become available, 8-azaguanine and guanine will also emerge as pleromers.

The usefulness of the concept of homotopy is, for the time being, more evident with respect to the amino acid sequence in proteins than to the nucleotide arrangement in nucleic acids. More and more instances of such homotopic relationships are being revealed as knowledge on specific sequences in proteins increases. As concerns the deoxyribonucleic acids, we are largely limited to the consideration of simple sequences, such as purine-pyrimidine-purine, or of small bunches of pyrimidine nucleotides. The development of the study of larger clusters, mentioned before, will doubtless contribute much to our knowledge. I have emphasized above a possible instance of homotopy, namely, the remarkable similarity in distribution, as solitary nucleotides, of the two 5-methylpyrimidines (thymine and 5-methylcytosine) in rye germ deoxyribonucleic acid³⁷.

At the present stage of our knowledge statements on the presence or absence of homotopy can be valid only in the most general statistical terms. Specific homotopic replacements may, of course, have taken place even where no all-embracing positional replacement can be affirmed. As concerns the replacement of thymine by a foreign analogue, discussed above in more detail, it ought to be understood that the observation of an ostensibly random replacement of thymine by 5-bromouracil does not exclude the possibility that the analogue may be involved in certain non-random preferences for neighbors. For instance, no infor-

* This regularity—in the absence of the other pairing regularities characteristic of deoxyribonucleic acids—may, formally, be taken to mean that in ribonucleic acids adenine can pair not only with uracil but also with guanine, and cytosine not only with guanine but also with uracil.

TABLE 36
EXAMPLES OF PLEROMERIC RELATIONSHIPS IN RIBONUCLEIC ACIDS

Source of RNA	Pleromers			Reference
	Analogue or natural satellite (mole %)	Normal pyrimidine nucleotide (mole %)	6-Am/6-K*	
Rat liver, soluble RNA	Pseudouridylic acid	Uridylic acid	0.95	54
<i>E. coli</i> , soluble RNA	Pseudouridylic acid	Uridylic acid	0.98**	55
<i>E. coli</i> , total RNA	5-Fluorouridylic acid	Uridylic acid	0.94	56

* Ratio of the molar sum of adenylic and cytidylic acids to that of guanylic, uridylic and pseudo- or 5-fluorouridylic acids.

** Ribothymidylic acid (1.1 mole %) included among the 6-keto nucleotides.

mation is available about the relative frequencies of pTpBp and pBpTp among the pyrimidine doublets or about the relative abundance of adenine–bromouracil–adenine as compared with that of adenine–thymine–adenine among the solitary pyrimidine units. Until much more is known it would be safer to assume that in the deoxyribonucleic acids pleromers are not necessarily homotopes, and homotopes not necessarily pleromers.

This brings me to the last topic of the present survey, namely, the neighbor problem. It was, I believe, first stated in the following terms¹⁰. “The frequency of methylcytosine being linked to guanine in a dinucleotide as compared with that of cytosine being so attached is twenty times as high in calf thymus nucleic acid, and thirteen times as high in the wheat germ preparation, than would have been expected for non-selective incorporation. The unexpectedly high proportion of 5-methyldeoxycytidylic acid that is linked to deoxyguanylic acid^{57, 58} makes one, in fact, wonder whether it is not the adjoining nucleotide rather than the one opposite that directs the incorporation.” This remarkable phenomenon—the high tendency of methylcytosine to be next to guanine—was again touched upon in our study on the sequence characteristics of the deoxyribonucleic acid of rye germ³⁷. Several analogous observations also were discussed there of which one may be mentioned. When, in this nucleic acid, the coupled dipyrimidine units are surveyed, it is found that methylcytosine has a greater tendency to be linked to either thymine or cytosine than the latter has to associate with itself or with thymine; and there is a definite bias in favor of methylcytosine–cytosine units.

Somewhat similar observations have also been made in the case of the ribonucleic acids in which the analysis of neighboring tendencies is easier because of the facility of obtaining any given nucleoside as either the 3'- or the 5'-phosphate. An examination of this type has, for instance, been made with ribonucleic acid preparations from different cellular portions of rat liver⁵⁹. A particularly instructive example of the neighbor principle is offered by the “soluble” ribonucleic acids of the cytoplasm concerned with the specific transfer of amino acids⁶⁰, in which the

terminal sequence adenine–cytosine–cytosine is required; a sequence that can apparently be enforced by a specific enzymic mechanism.

The manner in which a growing deoxypolynucleotide chain could guarantee the quality of its continuation is not yet a topic encouraging speculation*. Our present ideas about the enzymes taking part in the biosynthesis of specific polymers will probably have to undergo a stringent revision before the directed synthesis of complicated sequences will begin to be understood. For this, more than a biological form of Scotch tape is required. Certain sequential restrictions could, incidentally, be explained if, under circumstances, runs of more than one nucleotide were the immediate precursors of the polymer chain. We have, for instance, discussed the evidence, in rye germ nucleic acid, of the two dinucleotides cytosine–guanine and methylcytosine–guanine being able to substitute each other at random³⁷.

In considering the problem of the nucleotide sequence in deoxyribonucleic acids we have barely turned the corner. There is a long road before us; and we shall not see its end.

* "Whereof one cannot speak, thereof one must be silent." Wittgenstein, *Tractatus Logico-Philosophicus*.

REFERENCES

- ¹ E. CHARGAFF, *Experientia*, 6 (1950) 201. [See Chapter 1 of this book.]
- ² E. CHARGAFF, in *The Origin of Life on the Earth*, Intern. Union Biochem. Symposium Series, Vol. 1, Pergamon Press, London, 1959, p. 297. [See Chapter 7 of this book.]
- ³ E. CHARGAFF, *Trans. Intern. Congr. Biochem*, 4th Congr., London, 14 (1959) 21. [See Chapter 8 of this book.]
- ⁴ J. T. EDSALL, in H. NEURATH AND K. BAILEY (Eds.), *The Proteins*, Vol. 1, Academic Press, New York, 1953, p. 549.
- ⁵ C. L. SADRON, in E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vol. III, Academic Press, New York, 1960, p. 1.
- ⁶ M. D. MONTAGUE AND R. K. MORTON, *Nature*, 187 (1960) 916.
- ⁷ E. CHARGAFF, in E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vol. I, Academic Press, New York, 1955, p. 307.
- ⁸ E. CHARGAFF, C. F. CRAMPTON AND R. LIPSHITZ, *Nature*, 172 (1953) 289.
- ⁹ C. F. CRAMPTON, R. LIPSHITZ AND E. CHARGAFF, *J. Biol. Chem.*, 211 (1954) 125.
- ¹⁰ E. CHARGAFF, in *Fibrous Proteins and Their Biological Significance*, Symposia Soc. Exptl. Biol., 9 (1955) 32. [See Chapter 3 of this book.]
- ¹¹ J. D. WATSON AND F. H. C. CRICK, *Nature*, 171 (1953) 737.
- ¹² L. F. CAVALIERI, B. H. ROSENBERG AND J. F. DEUTSCH, *Biochem. Biophys. Research Commun.*, 1 (1959) 124.
- ¹³ G. BERNARDI, M. CHAMPAGNE AND C. SADRON, *Nature*, 188 (1960) 228.
- ¹⁴ C. F. CRAMPTON, R. LIPSHITZ AND E. CHARGAFF, *J. Biol. Chem.*, 206 (1954) 499.
- ¹⁵ E. CHARGAFF, H. M. SCHULMAN AND H. S. SHAPIRO, *Nature*, 180 (1957) 851.
- ¹⁶ E. CHARGAFF, *Harvey Lectures, 1956-1957*, 52 (1958) 57. [See Chapter 6 of this book.]
- ¹⁷ A. P. RYLE, F. SANGER, L. F. SMITH AND R. KITAI, *Biochem. J.*, 60 (1955) 541.
- ¹⁸ H. BROWN, F. SANGER AND R. KITAI, *Biochem. J.*, 60 (1955) 556.
- ¹⁹ D. ELSON AND E. CHARGAFF, *Biochim. Biophys. Acta*, 17 (1955) 367.
- ²⁰ G. R. WYATT, *Biochem. J.*, 48 (1951) 584.
- ²¹ H. S. SHAPIRO AND E. CHARGAFF, *Biochim. Biophys. Acta*, 26 (1957) 608.
- ²² E. CHARGAFF, *Cold Spring Harbor Symposia Quant. Biol.*, 12 (1947) 28.
- ²³ E. CHARGAFF, E. VISCHER, R. DONIGER, C. GREEN AND F. MISANI, *J. Biol. Chem.*, 177 (1949) 405.
- ²⁴ E. VISCHER, S. ZAMENHOF AND E. CHARGAFF, *J. Biol. Chem.*, 177 (1949) 429.
- ²⁵ M. E. HODES AND E. CHARGAFF, *Biochim. Biophys. Acta*, 22 (1956) 361.

- 26 C. TAMM, M. E. HODES AND E. CHARGAFF, *J. Biol. Chem.*, 195 (1952) 49.
- 27 C. TAMM AND E. CHARGAFF, *J. Biol. Chem.*, 203 (1953) 689.
- 28 C. TAMM, H. S. SHAPIRO, R. LIPSHITZ AND E. CHARGAFF, *J. Biol. Chem.*, 203 (1953) 673.
- 29 K. BURTON AND G. B. PETERSEN, *Biochem. J.*, 75 (1960) 17.
- 30 P. A. LEVENE AND W. A. JACOBS, *J. Biol. Chem.*, 12 (1912) 411.
- 31 S. J. THANNHAUSER AND B. OTTENSTEIN, *Z. physiol. Chem.*, 114 (1921) 39.
- 32 P. A. LEVENE, *J. Biol. Chem.*, 48 (1921) 119.
- 33 S. J. THANNHAUSER AND G. BLANCO, *Z. physiol. Chem.*, 161 (1926) 116.
- 34 C. A. DEKKER, A. M. MICHELSON AND A. R. TODD, *J. Chem. Soc.*, (1953) 947.
- 35 H. S. SHAPIRO AND E. CHARGAFF, *Biochim. Biophys. Acta*, 26 (1957) 596.
- 36 H. S. SHAPIRO AND E. CHARGAFF, *Biochim. Biophys. Acta*, 23 (1957) 451.
- 37 H. S. SHAPIRO AND E. CHARGAFF, *Biochim. Biophys. Acta*, 39 (1960) 68.
- 38 R. LIPSHITZ AND E. CHARGAFF, *Biochim. Biophys. Acta*, 19 (1956) 256.
- 39 H. S. SHAPIRO AND E. CHARGAFF, *Biochim. Biophys. Acta*, 39 (1960) 62.
- 40 T. TSUMITA AND E. CHARGAFF, *Biochim. Biophys. Acta*, 29 (1958) 568.
- 41 A. N. BELOZERSKY AND A. S. SPIRIN, in E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vol. III, Academic Press, New York, 1960, p. 147.
- 42 F. WEYGAND, A. WACKER AND H. DELLWEG, *Z. Naturforsch.*, 7b (1952) 19.
- 43 D. B. DUNN AND J. D. SMITH, *Nature*, 174 (1954) 305.
- 44 S. ZAMENHOF AND G. GRIBOFF, *Nature*, 174 (1954) 306.
- 45 D. B. DUNN AND J. D. SMITH, *Biochem. J.*, 67 (1957) 494.
- 46 R. M. LITMAN AND A. B. PARDEE, *Nature*, 178 (1956) 529.
- 47 S. BENZER AND E. FREESE, *Proc. Natl. Acad. Sci. U.S.*, 44 (1958) 112.
- 48 R. M. LITMAN AND A. B. PARDEE, *Biochim. Biophys. Acta*, 42 (1960) 117.
- 49 H. S. SHAPIRO AND E. CHARGAFF, *Nature*, 188 (1960) 62.
- 50 R. RUDNER, H. S. SHAPIRO AND E. CHARGAFF, *Nature*, 195 (1962) 143.
- 51 J. H. SPENCER AND E. CHARGAFF, *Biochim. Biophys. Acta*, 51 (1961) 209.
- 52 J. D. WATSON AND F. H. C. CRICK, *Nature*, 171 (1953) 964.
- 53 D. B. DUNN AND J. D. SMITH, *Biochem. J.*, 68 (1958) 627.
- 54 R. LIPSHITZ AND E. CHARGAFF, *Biochim. Biophys. Acta*, 42 (1960) 544.
- 55 D. B. DUNN, J. D. SMITH AND P. F. SPAHR, *J. Mol. Biol.*, 2 (1960) 113.
- 56 J. HOROWITZ AND E. CHARGAFF, *Nature*, 154 (1959) 1213.
- 57 J. D. SMITH AND R. MARKHAM, *Nature*, 170 (1952) 120.
- 58 R. L. SINSHEIMER, *J. Biol. Chem.*, 208 (1954) 445.
- 59 H. T. SHIGEURA AND E. CHARGAFF, *J. Biol. Chem.*, 233 (1958) 197.
- 60 M. B. HOAGLAND, in E. CHARGAFF AND J. N. DAVIDSON (Eds.), *The Nucleic Acids: Chemistry and Biology*, Vol. III, Academic Press, New York, 1960, p. 349.

CHAPTER 10

*A Few Remarks on Nucleic Acids, Decoding, and the Rest of the World **

*Men naar Contexten er bleven meningsløs, saa hjælper det ikke med store Overskuelser, saa gjør man bedst i at tage Talens enkelte Dele frem; naar Munden slaaer Sladder, saa hjælper det ikke at ville holde et sammenhængende Foredrag, men man gjør bedst i at tage hvert Ord for sig.***

KIERKEGAARD | EN LITERAIR ANMELDELSE | p. 115

1

One of our great men is reported to have said: When in doubt tell the truth. This is an excellent and not much followed precept. But our science has not been in doubt for quite some time. It has learned that it is profitable to outsmart nature.

Many of us, when we were young, thought that it was the task of the natural scientist to understand the ways of nature. But it has come about in our times that what man desires—or at any rate what some men desire—is to change these ways. It all started as a search for truth; but hundreds of thousands, at Hiroshima, at Nagasaki, paid with their lives for such lovable inquisitiveness. What an ivory, what a tower!

So I have learned to watch these clever fingers closely and not to believe the song of the sirens, even if they are sirens with a Ph.D. When they tell me: “As soon as we have mastered the genetic code, we shall know how to cure cancer”, my answer is: “Maybe; but before this, you will probably succeed in converting humanity into a bunch of mongoloid idiots, by putting just the right amount of base analogues into the DNA”.

* This is the text of a lecture given, in a much abbreviated form, at a symposium on “*Basic Problems in Neoplastic Disease*” held at Columbia University on March 12, 1962.

** But when the context has become meaningless, grandiose surveys are of no use; it is best to bring out the individual parts of the speech. When the mouth only babbles, it is of no use to undertake a coherent lecture; it is best to take each word in itself.

A few years ago, in 1953, I had the task of introducing a symposium on conjugated proteins, and I should like to cite a portion of my introduction.

"... Scientists, like little fishes, swim in schools. When we open one of our scientific journals these days, we find a very uneven distribution of topics. Some important fields are almost entirely neglected, others seem to explode into bursts of unbelievable mediocrity. Really valuable contributions in the fields most in vogue at present probably are just as scarce as those dealing with the stepchildren of present-day biochemistry. But not all disciplines make it so easy to call each mush a "homogenate", each soup a "partially purified extract", and to speak—when you have nothing whatever—of a "system". There is a real danger that our science may suffocate in its own excrements."

This, I fear, has come about. Prepackaged concepts have become widely used; a form of *collage* makes it easy to give a pleasingly abstract shape to what many of us were wont to consider the most concrete thing in the world. Instant Science: just take a teaspoonful and add hot air.

Conscience—that safety valve of human *hybris*—has no longer a place in science. A spurious, artificially ordered universe permits the bestial grinding out of facts, many of them actually pseudo-facts, micro-idols of an Easter Island of the mind. The *vox humana* may still be a stop on the organ; but where is it found in our scientific papers? We are being invited to dig deep holes to hide from the consequences of our atomic discoveries; but there will be no hole deep enough, once we have learned to manipulate the genetic material. Goethe's "Zauberlehrling" has become the prototype of us all. It will also be nice to watch the consequences of the attempt to tickle the Universe, as in the projected atomic explosions in outer space: Daedalus sending up Icarus to provoke the sun. These infinite spaces of which Pascal dreamed, is it not time that they came to the end of their infinity?

2

DNA now is a magic name, the philosopher's stone of our days, the quintessence of contemporary alchemy. Since I have been able to follow the development of this remarkable creature from

before the time when it went to nursery school—it was almost still in its baby carriage—a few words may serve to paint a backdrop. For decades, the sciences of biochemistry and of genetics had each gone their own way, stepchildren in turn of two older sciences: biochemistry of chemistry, genetics of biology. What to the one was the molecule, the gene was to the other. There was no bridge between the two, and they were on the whole quite happy. It may be remarked that also in the Tower of Babel the many nations got along quite nicely as long as they did not talk to each other; bedlam broke loose when togetherness set in. This point arrived for us between 1937 and 1944 when plant viruses were shown to be ribonucleoproteins, when the transforming principle in pneumococci was shown to be deoxyribonucleic acid. At about the same time—and I had my share of the guilt—the concept of “biological information” raised its head and began to sport a multicolored beard which has become ever more luxurious despite numerous applications of Occam’s razor. What must be remembered is that the entire idea of a chain of command, of a transmission belt of biological information could hardly have been formulated, had not the study of viruses made it possible to place a dateline on the start of development of a biological system. It is of course possible that we have been foolish in applying the principle of the unity of nature to such biological primitives and in extending the results from these studies to the entire realm of life. There are many regions in which the unity of nature cannot be invoked. By the way, it is quite astonishing that there is practically no evidence for the existence of such chains of biological information in higher forms of life, in multicellular communities, as have been derived from the study of a few viruses and micro-organisms. Even the loudly proclaimed “universal genetic code” rests, for the time being at least, mostly on highly brain-washed *E. coli*.

3

It can, however, not be denied that very impressive advances have been made in the last fifteen years in the chemistry, the

physics, and the biology of the nucleic acids, and even more so in what might be called their metabiology or, perhaps better, their parabiology: a field that could best be described as the borderline between genetics and mythology. Some of the allegorical paintings illustrating our views of the role of the nucleic acids compare favorably with the products of the minor masters of the rococo.

Perhaps the first advance that was made in the chemistry of the nucleic acids was the elaboration of sufficiently precise micro-methods that permitted the quantitative analysis of very small amounts. This was accomplished between 1946 and 1948 and it truly formed the basis of all later work; for it served to demonstrate the existence of a very large number of different nucleic acids and did away with the old tetranucleotide hypothesis. But before going on, I should pause for a moment and remember the names of those that came before, names that our time—long in bibliography, but short in memory—likes to forget: Miescher, Hoppe-Seyler, Piccard, Kossel, Altmann, Neumann, Jones, Steudel, Feulgen, P. A. Levene, Thannhauser, Hammarsten, Jorpes, Gerhard Schmidt, and Dische. Great work was done by them and by others, and even without the time-saving expedient of silly abbreviations.

The next thing that the chemical investigation of the nucleic acids accomplished was to demonstrate the species specificity of DNA, *i.e.*, to show that the DNA of a given species had a characteristic and unchanging nucleotide composition; a composition that appeared to be characteristic of the species, but not of any particular organ. And the third thing was the observation made in 1949 that there existed in all DNA specimens very peculiar base-pairing regularities. These observations, together with the X-ray studies four years later, had as a direct consequence the formulation of DNA as a regularly paired double structure.

Many of you will undoubtedly be aware of the numerous speculations about the GC content of DNA and its meaning for the physical and biological properties of the polymer, but few

will know that the distinction between the AT types and the GC types of DNA was first made here at Columbia, just as was the introduction of such much used terms as the hypochromicity, the denaturation, the heterogeneity of nucleic acids and many other things.

Now for some of the problems posed by the chemical structure of a single- or double-stranded polymer as it is assumed to exist in DNA; and much of what I have to say applies also to RNA. Glancing at an arbitrary segment from such a very highly polymerized macromolecule, we see that the nucleotides are linked 5' to 3', that their arrangement is arrhythmic, but that the complementariness regularities are maintained. The lowest estimate of molecular weight of such a polymer is about $6 \cdot 10^6$, except for a few small viruses; the highest estimate has no limit: the latest quotations on the market were around $180 \cdot 10^6$. Even with only ten thousand nucleotides per strand there are 10^{8000} isomers. We deal with four or five components; but the "alphabet" is not as meager as it may appear, for these few letters spell out "words" that are very long indeed: ten thousand letters or more. There could, therefore, be plenty of information for the unhappy cryptographer. But what is implied in such a statement as "DNA is the repository of biological information"? This is, of course, a very convenient slogan and may be taken as another example of the motorized anthropomorphism that disfigures much of our biological reasoning at present. The concept postulates the existence of a chain of information in the cell which is thought to be represented by a series of signals, as it were, emanating from a specific DNA and transmitted by a specific RNA messenger of matching composition to the unspecific ribonucleoproteins of the ribosomes, thus instructing the latter to proceed to the manufacture of specific proteins. A simple-minded chemist could, of course, ask himself whether this was more than a terminological scaffold concealing our ignorance, something of the order of an allegory, pretty but awfully vague.

But there can be little doubt: sequential specificity of high polymers of cellular origin is, as we have learned in the recent

past, one of the principal marks of the specificity of the cell. The approach to the problem of the sequence of nucleotides in a nucleic acid is, however, much more difficult than in the case of the proteins. The main reason is that we still lack a sufficient arsenal of different enzymes breaking predictable internucleotide bonds in a specific manner. So far, we have been compelled to rely mostly on the differences in stability of the purine and pyrimidine nucleosides when exposed to chemical degradation. It is indeed very regrettable that amazingly little work is being carried out at present on the search for new and specific nucleases.

The basic fact which is at the origin of all our sequence studies on DNA is the great difference in stability towards acid exhibited by the purine and the pyrimidine nucleosides. Whereas the glycosidic link between purine and sugar is cleaved with little difficulty, pyrimidine nucleosides are quite stable under these conditions. For this reason, it is relatively easy to deprive a DNA polymer chain of its purine complement without disturbing the arrangement of the pyrimidine nucleotides. We have called degradation products of this type, *i.e.*, chains which contain free sugar phosphates at the places where purines were, apurinic acid. Additional, slightly more energetic, degradation will now release the various pyrimidine clusters occurring in the original DNA. We refer to pyrimidines that find themselves interjacent between purines as "solitary" pyrimidine nucleotides; longer pyrimidine runs can be designated as "bunched", or more particularly as "couples", "triplets", etc. For purposes of such an analysis, which we have referred to as differential pyrimidine distribution analysis, it is not necessary to distinguish between the two strands of a double helix, if it occurs, since such a DNA polymer could be viewed as being joined at one end, or even at both ends, so as to produce one continuous polynucleotide. The analysis is based on the differential release of pyrimidine nucleoside 3',5'-diphosphates by β -elimination and hydrolysis.

The principal conclusions from a large series of such studies are: (1) The method permits the distinction of DNA specimens

that are indistinguishable by the analysis of their total purine and pyrimidine composition. Nucleic acids having identical total nucleotide contents, though derived from different organisms, differ widely in sequence characteristics. (2) A large proportion of the pyrimidines—and therefore also of the purines—occurs as tracts of three or more in a row. All-thymine sequences of five or even longer occur in many instances, but cytosine has so far not been seen in units larger than triplets. (3) The nucleotide arrangement appears not to be random. This could be taken as an indication that a DNA chain could indeed represent a meaningful text. But is a text without a reader still a text? “What is the sound of one hand clapping?” asked the Zen buddhists.

Which are, one might ask, some of the immediate problems in the chemistry of the nucleic acids? First, a better understanding of what is meant by a DNA molecule. Is the apparent heterogeneity of DNA ostensible or real? Is there one giant molecule of DNA per chromosome—of a molecular weight of hundreds of millions—or are there many, different both chemically and functionally?

Secondly, if we really deal with giant chains, do they consist only of polynucleotides or are many tracts, each consisting of several hundreds of nucleotides, held together by bridges that are not of nucleotide character? The fact that DNA is depolymerized by alkali to smaller polynucleotides, but still of respectable size, would seem to indicate the second possibility. I have been interested in this problem for years, but with inconspicuous success. There is an insuperable difficulty in demonstrating, among hundreds or thousands of components of one kind, one of another. It is very hard to distinguish between a trace constituent and a trace contamination. One could think of many types of possible bridging compounds. A derivative of a hydroxyamino acid, such as phosphoserine, or a bifunctional amino acid, *e.g.*, glutamic acid, could form a diester, or an oligopeptide could do the same. If something of the order of glycerophosphate, ribophosphate, or even a ribonucleotide, were inserted into the chain at infrequent, but possibly predetermined, intervals, regions of reactivity dif-

ferent from the bulk of the polymer would be created. Alternatively, it would be sufficient for a purine or pyrimidine to be missing occasionally, so that at these places a deoxysugar phosphodiester occurred; in this manner less stable links would be introduced at which the chain could cleave *via* elimination or cyclization. The problem of the existence of discontinuities in the DNA chains is not without general interest, since these could serve to explain how such a polymer, and even a circular structure, could rupture at fixed and predictable positions, as has sometimes been assumed in bacterial and phage genetics.

And thirdly, and perhaps most importantly, if there really is a transcription belt starting from DNA, as I have mentioned before, what are the chemical mechanisms that are at play? How are the precursors selected and held in place, how are nearest-neighbor preferences exercised? There is a real problem in the question by what forces the precursors of a polymer—nucleotides or amino acids—are held onto the “template” before they are polymerized enzymically. Such silly metaphors as “the zipper” will not do. Anyone who has had trouble with this useful contraption knows how important the rigid alignment of “precursors” is for its proper functioning.

4

This brings me to the last topic I wanted to mention. Majority opinion—which means, as usual, a small minority with a strong voice—postulates that DNA or RNA carries a code which is, in the last resort, translated into the specific amino acid sequence of a protein. This code is said to be non-overlapping, composed of nucleotide triplets that are read from a fixed starting point, and “degenerate”, *i.e.*, several signs may have the same meaning. It is also thought to be universal.

“CODE - 2. Any system of principles or rules; as, a *code* of ethics. 3. A system of signals for communication by telegraph, flags, etc.; also a system of words or other symbols arbitrarily used to represent words; as, a secret *code*.” (*Webster's New Collegiate Dictionary*). — Number 2, I decided, obviously does

not apply, so it must be No. 3. But are we concerned with words here, and is arbitrariness what we would ascribe to nature? The understanding of words, the translation of ciphers, even the reception of signals, they all require the intervention of intelligence, either directly or indirectly. I had been interested in this problem for quite a while, long before I read Teilhard de Chardin from whose book (*Le Phénomène Humain*) the old saying of Tertullian seemed to arise in a new and disturbing version: *Animal naturaliter christianum*. He found consciousness, *i.e.*, mind, everywhere, though in different degrees, in the mineral and the protein, in the virus and in man. Now, this is a dangerous, a slippery path; and I am afraid to take even the first step on it. For if I fear anything, it is the mystic with the slide rule. Poets look better without their laboratory coats. The vapor pressure of *pneuma* should remain unmeasured. To study the effect of mental concentration on reaction velocity I leave to the next century.

In any event, when the word "code" is used in the context of cellular specificity, most people will know what it signifies in a loose sense, though many may be baffled by its concrete meaning and even more by the manner in which it is supposed to operate. For this reason, we were all elated—though perhaps with different degrees of fervor—by the announcement, a few months ago, that the amino acid code had been broken through the discovery that several uridylic acids in a row, perhaps three or more, coded for phenylalanine. These findings were soon confirmed and extended in other laboratories, so that it would now seem that the coding units for all amino acids are known, though there still are some uncertainties, especially with regard to the sequence within the triplets. There were, of course, some very baffling circumstances, *e.g.*, the unexpected preponderance of uracil in all coding units and also the observation that poly-U was enormously more effective in incorporating phenylalanine than were any of the other coding units for their respective amino acids.

We had a very direct interest in all these questions, for we were determining the frequency of the all-pyrimidine runs in various species of DNA; and there should have been a direct

relationship between their abundance—and also between the DNA type, AT or GC—and the coding efficiency of a DNA.

But before engaging in much work, I thought the situation over. And a short time ago I underwent a sort of conversion which lasted for one week. I decided to believe everything—but everything—that I read in the literature. I swallowed all biochemical snapshots, preliminary communications, all press releases that punctuated the, more or less, friendly competition between different laboratories—in other words, the entire sour grapevine. I followed the international spelling bee with open-mouthed admiration. I became what you might call a molecular fundamentalist. And this is what I encountered during my brief pilgrimage through the bible-belt of molecular biology. The dogma stated: DNA makes “messenger” RNA, makes protein. “Fine”, I said; but though temporarily a True Believer, I had not stopped being a chemist. So I reasoned: “Messenger” RNA is said to be a complete image of DNA, mirroring its base-pairing and other regularities; A = U; G = C, and so on. The amino acid composition of many proteins is known precisely, and in many instances also the amino acid sequence, thanks to the excellent work of the past few years. The amino acid code has been published, at least in the newspapers. So let us reconstruct the messenger coding for a given protein. This scheme was the more inviting, since the determination of the nucleotide sequence of an RNA by direct means is beset with very great difficulties.

TABLE 37
BOVINE PANCREATIC RIBONUCLEASE
Number of amino acids: 124

<i>Calculated composition of “messenger” RNA</i>	<i>As mole %</i>	<i>Ratios</i>	
A ₆₃	A 16.9	Pu/Py	0.49
G ₅₉	G 15.9	6-Am/6-K	0.60
C ₇₇	C 20.7	A/U	0.36
U ₁₇₃	U 46.5	G/C	0.77
		(A + U)/(G + C)	1.7

First, I chose pancreatic ribonuclease, and the results are shown in Table 37.

You will notice that there are many surprising features as regards this fictitious RNA. It does not look at all like anything that we are familiar with. Almost one half consists of uridylic acid; the pyrimidines amount to two thirds of the whole; the ratio of 6-amino to 6-keto nucleotides is only 0.6. One could not be farther away from an image of DNA. It turned out that one could; for I chose as my next example a protein closely associated with DNA in a sperm cell of which it constitutes a considerable proportion.

I took salmine, the protamine that, next to DNA, is the principal component of salmon sperm (Table 38).

This reconstructed "messenger" looks even more unreal. It contains only one adenylic acid in 174 nucleotides. Even if

TABLE 38

SALMINE

Number of amino acids: 58

<i>Calculated composition of "messenger" RNA</i>	<i>As mole %</i>	<i>Ratios</i>	
A ₁	A 0.6	Pu/Py	0.40
G ₄₉	G 28.2	6-Am/6-K	0.49
C ₅₆	C 32.2	A/U	0.01
U ₆₈	U 39.1	G/C	0.87
		(A + U)/(G + C)	0.66

TABLE 39

SALMON DNA

<i>Mole %</i>	<i>Ratios</i>	
A 29.7	Pu/Py	1.02
G 20.8	6-Am/6-K	1.00
C 20.4	A/T	1.02
T 29.1	G/C	1.02
	(A + T)/(G + C)	1.43

one assumed that the "messenger" is produced in the form of two, presumably antiparallel, strands as an image of the particular segment of DNA, of which only one carries the "coded" information, a sequence of salmon DNA which in a tract of 174 nucleotides contained only one adenylic acid looks most unreal on many grounds. For instance, salmon DNA is of a quite marked AT type (Table 39).

You will notice that in salmon DNA almost every third nucleotide has a good chance of being adenylic acid. How probable is it then that this nucleotide occurs only once in such a long stretch? Moreover, even a glance at the published "code" will convince us that there can be very few proteins that would demand a less implausible messenger. We are, hence, left with the assumption that each double segment of DNA can specify, *via* two RNA chains, two different amino acid sequences of which only one is realized. But what does the other half of "messenger" do and what is its fate? Surely, nobody would like to put forward the inane slogan: One gene, two enzymes.

Another explanation, but not a very good one, of these difficulties may be seen by some in the assumption that the code is "degenerate", *i.e.*, more than one triplet may specify one amino acid. It should, however, be pointed out that if the "degeneracy" took the form of having one triplet coding for a given amino acid in one species, but another coding unit for the same amino acid in another species, the proclaimed universality of the code would vanish. And then there is, as always, still another possibility, namely, that the whole thing is wrong.

In any event, when I saw all this I was very baffled. I concluded that what I had tried to do will not be a shortcut towards a solution of the problem of nucleotide sequence in RNA and DNA. I went home and lost my faith.

5

Most scientists like to believe that we live in a cosmos, *i.e.*, in a universe having order, harmony, purpose. Many feel that in this world all hangs together, though few will presume to explain how

it does so. "Pound St. Paul's Church into atoms, and consider any single atom; it is, to be sure, good for nothing: but, put all these atoms together, and you have St. Paul's Church." (Johnson in *Boswell's Life*, July 20, 1763). You can see, Samuel Johnson was what you might call a Premature Molecular Biologist; otherwise, he could have considered the possibility that reassembly may yield Liverpool Station. Whether his limpid eighteenth-century mind would have enjoyed the gnostic stew that is being served up at present remains, however, questionable.

We have become insensible to the incredible amount of noise and tribal incantations that accompany scientific research in our days. "*Tolle, lege!*" said a voice to St. Augustine. But if in our time he had, following this injunction, picked up and read one of our foremost picture magazines devoted to illustrating the more appealing and intimate sides of our public personalities, and therefore hard up for copy, this is what he would have seen as a headline in a recent issue: "Will man discover 'the molecular structure of God'?" If anything can turn the stomach of a saint, this is it. Has our time forgotten how heavy a word can be?

CHAPTER 11

Amphisbaena

...et tu in te manes, nos autem in experimentis volvitur?

AUGUSTINUS | CONFESSIONES | IV, 5

*'Tis all in peeces, all cohaerence gone;
All just supply, and all Relation.*

DONNE | AN ANATOMIE OF THE WORLD
THE FIRST ANNIVERSARY, 213-4

Nay, so devoted are we to this principle, and at the same time so curiously mechanical, that a new trade, specially grounded on it, has arisen among us, under the name of "Codification", or code-making in the abstract; whereby any people, for a reasonable consideration, may be accommodated with a patent code; — more easily than curious individuals with patent breeches, for the people does not need to be measured first.

CARLYLE | SIGNS OF THE TIMES (1829)

Two men sit on a bench, in August 1961, an Old Chemist (O) and a Young Molecular Biologist (Y).

O:

Now that I have you here for a few minutes of quiet talk, I shall start by saying: The cell is not a machine.

Y:

But what kind of a start would that be? Do you want me to agree? This would be surely stupid, for there are all sorts of machines, and you have probably never heard of the theory of automata. I should much rather turn your saying around and offer: A machine is a cell.

O:

We are then already in the middle of model building, the favorite occupation of modern biophysics. It is all done in front of mir-

rors, with wire and plastic, glue and papier-mâché; the knowledge of a child combined with the naïveté of the grown-up.

Y:

But really, why are you so much against machines?

O:

Of course, I am not; some of my best friends are machines. But I am very much against a strictly mechanomorphic view of living nature. A machine is a deterministic construction; someone—an intelligence—has had to make it; and even if it could be “programmed” to make itself, who did the programming? A self-reproducing machine that was not built by a primordial engineer is an abomination thrown up by turbulent and sick times which, in chiliastic dreams mixing superpower and impotence, have created a mythology of a most maculate conception. I shall not formulate the secret of life sitting on this bench on a summer day; but I can say that we shall not have a satisfactory theory of biology before we have learned to combine, in one concept, the dialectic play between determinism and accident—a sort of random nonrandomness—which seems to characterize the living and reproducing cell.

Y:

It would seem to me that you are simply an angry old man preaching some sort of dialectical biochemistry in which the observer is walking simultaneously on both sides of the street in opposite directions. One could call this also schizophrenic.

O:

Well, as to being an angry old man, there is plenty to be angry about, and it makes more sense for an old man to be angry than for a young one; like most things, anger must be earned. King Lear was an angry old man. We are no longer used to passion in the natural sciences; it has been replaced by ambition. Our young geniuses are passionately ambitious instead of being pas-

sionately passionate; and it has become very difficult to distinguish between what is an ardent search for truth and what is a vigorous promotion campaign. What started as an adventure of the highest has become the survival of the slickest or the quickest. "Cloak and dagger" has changed to "cloak and suit". We now have DNA tycoons and others have "made a killing" in RNA. A generation of scientific quiz kids knowing the answer to everything. But to throw pearls before young molecular biologists is not the purpose of this talk. Let's return to the beginning and I will ask you "What is molecular biology?" Now, if physics is the science of the states, and chemistry that of the conversions, of matter and biology comprises the application of their laws to animate nature, what could be meant by molecular biology? You can, of course, apply any adjective to any noun, but the results are most often bizarre.

Y:

Here you go again. I could, of course, be witty, as you think you are, and say that molecular biology is what is published in the journal carrying that name. All academic gowns carry, strictly speaking, the color of their respective payrolls—

O:

Here I interrupt already. Your jocular definition cannot hold, since the irregular rate of appearance of this admirable journal is vastly inferior to that at which molecular biologists are produced nowadays. My definition, incidentally, would be that molecular biology is essentially the practice of biochemistry without a license.

Y:

This is, as you must know, a completely frivolous definition. There is naturally much more to molecular biology. In order to prevent you from going on endlessly, I shall agree that the naming of a new science—or a new name for an old one—has also its practical reasons. Symposia, congresses, new journals, more money to be had more easily . . .

O:

And the feeling to be a pioneer at no extra cost. And how the publishers love these new handles by which to extract the old money. In fact, this is not the end: soon we shall have molecular sociology, molecular history, and a little later perhaps molecular theology. The fragmentation of the sciences proceeds *via* adjectives.

Y:

I believe, even among your age group you are a horrible exception.

O:

Very likely. In a time in which everything that is new is true, the older the foggy the more he must galvanize himself into being able to participate in the exuberant initiation dances that greet each new molecule before it is replaced by an even newer one. But I must say, you can't escape senility by trying to become a juvenile delinquent.

Y:

I shall ignore this. What I said a moment ago about the advantages of having a new science was not meant satirically. This is the way sciences grow in our day; they have become a mass movement, a corporation in which the majority of the population carry stock. Their development must be studied by the methods of sociology.

O:

Another stew of a science.

Y:

Let's slaughter one at a time. To return to our subject—if we still have one—this is, I believe, the way our new science came about. It started with the recognition of the importance of macromolecules in biology, of their chemically distinct and pre-

cisely describable structures, and of the specific functions which very often could be clearly assigned to them. Take the enzymes . . .

O:

I will take them. But who did the assigning? You isolate a protein, you ask it a very limited number of questions by bringing it together with a few substances. If it happens to react with one or the other, you call them "substrates" and you proceed to assign a specific function to this particular protein. Did you notice the man that passed us a moment ago? He limped; and you would rightly say that this was because he had one shorter leg. But how could I refute someone who claimed that this man had one leg too short in order to limp? You say that this enzyme is present in the cell, in order to perform this reaction. *Hic Rhodus, hic salta!* But maybe this is not Rhodes or maybe he does not jump, or maybe he jumps somewhere else. How sure can you be, for instance, that all the various phosphatases that can be fished out of a cell mush really act as such in the living cell, that this is their "function"? Ours—and yours—still is a *post mortem* science; we are forced to destroy the overriding, the overpowering category of life.

Y:

Don't tell me you are a vitalist.

O:

Of course, not. But a discussion between us about the meaning of life would be stupid: I am too rigid and you are too flexible. All I can say is: Life is what's lost in the test tube. Better tell me what you think about what I said concerning the assigning of functions.

Y:

I think there are such enzymes and such. Many have the functions that we have given them, others may not be called upon to act

under normal circumstances; they may be a memory that the cell has retained of past events. But surely even you cannot deny that, for instance, the energy relationships of the cell and their enzymic basis are well understood and are the best example of the concept of the unity of biochemistry?

O:

This I will not deny, though I cannot say that I am particularly fond of this concept of unity. It has done a lot of mischief. Life, as we know it, seems to be characterized by two antithetical principles: unity and diversity. And it is very difficult to decide, in a concrete case, with which of these principles we are dealing. The unity of nature is usually laboriously pieced together by combining, in a spuriously serial or cyclic fashion, snips and bits taken from the most widely distant organisms. It is a sort of biomontage. The fact that we all live and die in one and the same world should not conceal another fact: that we are all different. Even life is only one, and a minor, form of nature: a tiny foam on the crystals of the earth. For an old man who is in love with what goes on around him and with the multiformity of its appearances there can be only one battle cry: *Vive la différence!* Is it really so important that both torturer and victim enjoyed the same hearty meal and digested it in the same manner? Besides, similar mechanisms do not always foreshadow similar functions. The combustion oven designed by Liebig and the combustion ovens constructed in Auschwitz, though based on the same scientific principles, can hardly be taken as a proof of the unity of nature. But your little speech about the processes of parturition that resulted in the forceps delivery of molecular biology stopped at the enzymes. Perhaps you go on.

Y:

What you like to call my little speech did not stop, it was stopped. In any event, the recognition that polypeptides of high molecular weight could show great specificity as enzymes in metabolic reactions and as antigens or antibodies in immunology

led to many studies of their physical and chemical structures. Ultracentrifugation, X-ray diffraction, electrophoresis, electron microscopy, the many wonderful separation methods carrying the name of chromatography: all these were instrumental in bringing about these exciting times. It became possible not only to analyze proteins completely but even to determine their amino acid sequence in many cases. The triumphs of structure determination were so great that everywhere young people were flocking to the banner of the helical nature of things.

O:

You are spiraling a little too fast for me. Even Lucretius had to stop for breath. Your dithyrambic tone reminds me that not long ago I have heard a gushing voice on the radio speaking of "DNA, this miracle monocule".

Y:

What of it? Some people have difficulty pronouncing "molecule"; but this will change. Just as there were times when people found it easy to say "transubstantiation". As to DNA, I shall soon come to it. Let me continue. The advances I mentioned a minute ago were, of course, not the only ones. Even before they had come to fruition, studies had begun on other fronts. Certain viruses had been purified sufficiently so as to warrant their chemical examination and had been recognized as nucleoproteins. First came the plant viruses which were found to contain RNA, as do also most of the animal viruses. The bacteriophages, on the other hand, contain mostly DNA. But perhaps the most exciting thing that happened was the discovery that microbial transformation—a phenomenon known for quite some time—was due to specific forms of DNA. This had the effect of placing DNA very near to what the geneticists at that time were still calling the gene.

O:

If you don't mind let us still stick to this hoary term. I know, the packaging has been changed lately; but since it still is the same

old thing, why not conserve the convenient brand name? After all, if you want to describe a syndrome you must first give it a name. The one advantage, for instance, that I can see in using the term "molecular biology" is that it puts nearly all that is unknown in biochemistry into one convenient corner. But I wanted to say a word about transforming DNA. I well remember the great excitement with which I followed the first discoveries concerning specific pneumococcal deoxyribonucleic acid. Then we still spoke of "desoxyribonucleic". In fact, this was probably the main reason for my becoming interested in the nucleic acids at about that time, 1944 or 1945. But there has been a considerable period since then and not so very much has happened in this field; and I must confess that my misgivings about the justification of expanding these few observations on micro-organisms to the entire realm of life have been growing at an ever increasing rate. Generalization has its definite uses in science; without it we should all soon be without a job. But at the same time, there is a great danger of its gliding into glibness. It was in 1889 that the great Swiss historian Jacob Burckhardt wrote a letter to a friend in which he warned of the oncoming of what he called "*les terribles simplificateurs*". Just as the locusts, once they are through with a field, have simplified it horribly, could we not say that this is also true of some of the great generalizations in biology? Color and variety, the pulsations of accident and fate, the tremendous urges and instincts, the pendulum of birth and death: all have disappeared and we are left with what I once called "a plantation of match sticks". So, when I listen to the arguments that microbial transformation proves the gene character of DNA, I must ask: is this discovery one of the features of the unity of nature or one of the facets of its diversity? That's where the dialectics has to come in to which we referred at the beginning of our little disputation.

Y:

I can see, you want to keep your cake and sell it. You are an intolerable mystic, and I have had the feeling that, while you

speak of nature, there should be a slight music of harps. Why do you keep plucking the feathers out of the goose that lays golden eggs for all of us? After all, you are one of the classics in this field.

O:

Thank you, I should rather not be. In your definition, a classic in science is a man who does not have to be quoted any longer. For the pickpocket, the man with the widest pockets is a classic. And it has been remarked that Banquo is seldom quoted in Macbeth's papers.

Y:

All right, then I shall call you a Cassandra in long pants. But my story was not yet entirely finished. I had been talking about transforming DNA; and if you had not broken in, I should have continued by mentioning other instances in which nucleic acids were assigned direct roles in the determination of hereditary properties. For instance, there are the bacteriophages attacking *E. coli* which seem to do it by injecting their specific DNA into the bacterial cell. This is enough to set off an entire chain of events terminating in the production of many phage particles and the rupture of the host cell. And then we have the plant viruses; they contain specific types of RNA which are infective; that means, the nucleic acid itself, when applied to the plant, can give rise to the formation of innumerable complete virus particles. These are all definite and specific molecules exerting definite and specific biological effects; and here you have the quintessence of our new science, molecular biology. But this is not all.

O:

I am sure, it is not. Even to sell soap nowadays, you need an *a cappella* choir. What baffles an old-fashioned chemist in all this is what has become of the chemical concept of a molecule. I have heard people of a still older generation claim that this concept ends with the applicability of Avogadro's law. But never mind.

In any event, it must have been a pleasant surprise for some biologists to learn that at last they were dealing with molecules. I should have thought that this was what they had been doing all the time. It reminds me a bit of Monsieur Jourdain in Molière's *Bourgeois Gentilhomme* who is astonished to hear that what he has been speaking all his life is called "prose".

Y:

Surely, you must have heard the term "molecular disease"?

O:

Alas, I have; and for a time I thought that this was a disease to which molecules were particularly prone, a sort of molecular measles. But I soon learned that this was another symptom of the general sloganification of science, everything—like in a poor cartoon—having a pithy label hanging from the mouth. Some of these slogans may have been convenient or useful at one time, such as "energy-rich phosphate bonds", "dynamic state of body constituents"; of others, for instance, "messenger RNA", I am less sure. But there are so many of them, and they are so glittering, so glib! The more I hear them, the less I enjoy them. The peculiar relationship between names and understanding has often been discussed, best perhaps by Mephistopheles.

Y:

Great concepts require great names.

O:

Or perhaps great names can substitute for great concepts. But I believe, you had not yet finished.

Y:

This is correct, for I now have to introduce the crowning concept, namely, "biological information".

O:

Do you mean to say that life itself has now acquired a press agent?

Y:

Of course not; but there has arisen all over the world a group of young, militant and successful practitioners—evangelists, you might call them—who are spreading the new knowledge with devotion and perseverance. All seems to fit.

O:

Or at least you disregard what doesn't. Such must have been the atmosphere when Phlogiston went strong: all seemed to fit until a little balance was brought into play. Even the great Ockham with his immortal razor would no longer succeed in shaving the multicolored beard that is now sprouting all over the beautiful face of biochemistry. Opening one issue of a journal, I am assailed by an abysmal cackle of terms, ill-fitting, yet strident, garish and banal. The shockate, the grindate and the sonicate, the suicide and the abortion, the fingerprints and the hot spots, the repressors and the co-repressors, the feedbacks, the pools and the templates, the regulators, the operators and the operons; and floating over the entire allegorical cesspool, the mysterious messengers, angelic or diabolic in their evanescent everywhere-ness. Have we really arrived at the stage of non-objective biochemistry, of molecular action painting?

Y:

You have left out the hybrids and many other things; but I am glad to see that you are fairly familiar at least with our nomenclature. Let me say, however, that to make a scientific revolution one must break many eggheads. Your objections are ineffective and they will not count in the end. Biological or, if you prefer, genetic information is an extremely important and useful concept which even a Shakespearean fool cannot laugh off. Before I go on, I should like to ask you a question. When trying to

understand the life of the cell and the functions of its individual parts, don't you believe in the division of labor?

O:

Frankly, I don't; at least not in the crudely mechanomorphic way—the wheels and the gears and the levers—in which this is usually done. Even if Nature were one gigantic servo-mechanism, I am afraid the beards of the cyberneticists, entrusted with servicing it, would get into the way of the feedbacks. In the living cell there must be a way in which quantity—or, better, density or compression—regulated on an as yet undescribable time scale, becomes a new and unique quality, namely, that of life.

Y:

I am afraid, you are, after all, a vitalist. Returning now to this business of information and disregarding all you have said, this is what it amounts to. We do believe in the division of labor in the cell, with each part, yes, with each molecule, having a definite and recognizable function. And we believe in the existence of a strict hierarchy.

O:

I know, I know. Mix anything with everything in the right proportions and the resulting purée will say: Papa! But tell me, since you mention the hierarchy of the cell, reading the recent literature I get the impression that the cell is a society of slaves that have no master.

Y:

Not at all. In the beginning was DNA . . .

O:

I hear the start of a new apocryphal gospel with DNA as the logos of our times.

Y:

Is it possible that you don't believe in DNA?

O:

If I don't believe that the moon is made of green cheese, does this necessarily mean that I don't believe in the moon?

Y:

Anyway, DNA is the genetic material which in the last resort is responsible for the maintenance and the transmission of the hereditary properties of the cell. We know about the mechanism of its replication which is an ingenious deduction from the generally accepted structural model of DNA: a double helix composed of two intertwined polynucleotide chains. After the separation of these strands which is very easy to explain, as it offers no thermodynamic difficulties and can be carried out with a model costing less than a dollar, each strand now proceeds to the production of its counterpart.

O:

How general is general? And without wanting to offend your sense of economy, thermodynamical or otherwise, may I ask whether you mean this beautiful scheme to apply to meiosis as well as to mitosis?

Y:

I am not interested in diploids.

O:

I wish, your parents had felt the same way. And would you say that during the splitting of a chromosome the other constituents, the proteins, the lipids, undergo similar processes, or do they simply follow the leader out of family affection?

Y:

Let's not waste our time with trivialities. Who cares about lipids anyway? The smart cookies leave them alone. I shall give you here a purely formal scheme which you can take or leave; but you would be well advised to take it.

O:

I know, scientific dogmas are phagocytes which eat only what is good for them. They cannot be refuted or dethroned; but they vanish eventually owing to the fickleness of subsequent generations who lose interest in them. In fact, the more absurd a hypothesis is, the stronger must be the belief in it.

Y:

I continue. DNA is, as I said, the primary genetic determinant carrying a code, as yet unbroken, through which in the last resort the composition of RNA and of the proteins is specified. Before this code can be expressed, the two strands composing the double helix of the DNA must be separated, unwound, perhaps enzymatically. I believe, it was you who called this hypothetical enzyme an "unscrewase". You will observe that, because of the complementary structure of the two strands, the information stored in either one really is sufficient.

O:

Have the two strands been shown to exist, let alone to have a complementary structure?

Y:

Well, yes and no. But you should be the last to ask such a question.

O:

Sometimes I wake up in the dark of the night and I begin to think of all these claims and discoveries and models, of all this molecular prestidigitation; and I ask myself: Is this all a confidence game? They are all so brilliant; why are they so shallow? Why did the manna of the heavens turn into porridge? Why does the liquidation of a science begin at the top; why do its greatest triumphs turn into its worst disasters?

Y:

I don't believe you expect an answer. All you have to do is to take a spectra of a DNA before and after heating.

O:

Now I can see that you are a modern young scientist. The use of these false singulars, a spectra, a media, a bacteria, a phenomena or even a phenomenon, etc., is this not the *summa cum laude* of the new generation?

Y:

I have never found Latin useful in my work. If you, too, had studied more mathematics and physics, instead of wasting your time, you would have accomplished more.

O:

This may very well be so. I have felt for some time that your Ph. D. should be spelled pH D.

Y:

In any event, this is how it works. There is no doubt at all, the direction is: DNA to RNA to protein; but never backwards. The genetic signals emanating from the DNA go first to a special kind of RNA which turns over very rapidly, never amounts to more than a few per cent of the total RNA of the cell, and mimics, in its composition, that of the DNA. Since it carries the message contained in the DNA, we call it the messenger RNA.

O:

This small and short-lived Corps of Commissionnaires, has anyone seen it?

Y:

What do you mean by "seen it"? There is plenty of indirect evidence that such an RNA is formed.

O:

No, I mean the message, has anyone shown that it exists? Is it not possible that the entire imposing terminological scaffold is nothing but a suitcase for the emperor's new clothes? Is it not

possible that there is no message, no messenger, that the entire question is asked, and therefore answered, wrongly?

Y:

If you deny that there is such a thing as "information", you will still have to get the same fellow, but call it by a different name.

O:

You want to say what is usually said when one has no answer, that this is merely semantics?

Y:

Yes. Incidentally, messenger RNA is not the only name; others have referred to it as informational RNA or pulse RNA. The men that really discovered the stuff gave it no name and will therefore be rightly forgotten.

O:

This is true, never be a pioneer; not who comes first counts in science, but who comes last. But let us stay with these charming conceits for a little while and try to think things through. I know, in our days what cannot be done indirectly, isn't being done at all any more. But chemistry is, after all, the science of substances; and if all this is true there must be a substantial basis to it. You said that it all starts with DNA. This DNA, I take it, exists in the cell as a double-stranded helix, as what people with a brilliant gift for vulgarization have called "the coil of life". The two strands run in opposite directions with respect to their terminal phosphates and have a complementary nucleotide sequence, so that purines can pair with pyrimidines, adenine with thymine, and guanine with cytosine. And you might say that this scheme was confirmed before it was devised, since this had been shown to be true of the composition of DNA. It has been forgotten completely that the specific base-pairing in nucleic acids was discovered in an old-fashioned chemical laboratory.

Y:

Well, I happen to remember, but who cares? What did these people whom you call old-fashioned do to make their findings known?

O:

They published them.

Y:

Published? Are you being jocular? Is this what you call aggressive scholarship? Did they send out mimeographed copies of their papers long before publication? Did they found base-pairing clubs? Did they distribute neckties with suitable emblems?

O:

But who would want to spend his life dancing a minuet before assembled science reporters? There still are people who do not wish to join the noise boys; they have other concerns. When you start on something new, you are all alone and it is so terribly dark; and then, suddenly, you may come face to face with the blinding whiteness of reality. There is nothing more exquisite, nothing rarer in the world. Afterwards you have a choice: You stay in the laboratory, hoping that it will happen again; but it seldom does. Or you begin to travel through the country giving minstrel shows.

Y:

Anyway, even if the composition of DNA had been found entirely different, we should have doubted the analysis, not our concept.

O:

I can see, you are a True Believer. It is with such deductions that the road to the paradise of scientists is paved. But really, I did not wish now to attack what one is pleased to call the "central dogma", for I know that the mythopoeic urge of humanity does not stop at the door of the laboratory. But let me continue,

and I shall assume that the existence of two complementary strands, each having two ends, a head and a tail as it were, has been proved.

Y:

Well, if it hasn't, it will be.

O:

This DNA, as I understand it, has two functions: it has to make itself—that is easy: unscrew, assemble, polymerize, recombine. This has now become a so-called project for the so-called Science Fairs of our so-called high-school boys. But it also has to make the rest of the cell; and this is not so easy. For in this DNA must be contained the quintessence of all this tremendous life of our earth, the flagellum and the spirillum, but also the brain that invented the St. Matthew Passion. “Dinanzi a me non fuor cose create . . .”

Y:

Don't talk Latin at me.

O:

Never mind. What I wanted to point out is that in our days DNA plays the role of a self-replicating philosopher's stone. First we have an extremely vicious circle: DNA makes DNA makes DNA, etc.; a dreary, tragic desert, an Yves Tanguy landscape. But simultaneously—and by what delegation of functions even you could not tell me—the Dr. Jekyll side of DNA gets into the act and makes A and B and C. I am always being told that all biological information resides, in the last resort, in DNA. But organisms contain many types of specific molecules, quite apart from the proteins and the nucleic acids: the mysterious conjugated proteins, some of the lipids, cell wall substances, blood group substances, specific polysaccharides.

Y:

These polysaccharides contain no information.

O:

How do you know? Did you talk with them lately? But let me return to our argument. When DNA gives play to the friendlier part of its nature, the first thing it does, I am told, is to preside over the manufacture of messenger RNA. This RNA is said to show the composition regularities of the entire DNA. If the latter really consists of two complementary strands, this must mean that two complementary RNA strands, or a mixture of shorter pieces amounting in the aggregate to two such strands, have been made. To put it more concretely, a triple uracil in one messenger polynucleotide would have to correspond to a triple adenine in the complementary RNA structure. Since a ribosome—"they also serve who only stand and wait"—can make a protein only when it is "programmed" by a messenger RNA which on the whole cannot be a double strand, as it must be able to engage in specific hydrogen-bonding, the conclusion would have to be that any given section of a DNA dyad should give rise to two entirely different proteins. Would you consequently subscribe to the revised slogan "one gene, two enzymes"?

Y:

Well, I don't know. There are hundred ways out.

O:

But should a scientist behave like a tracked cockroach? I know, almost anything you can write on a piece of paper will eventually be realized in a so-called system and then it will form a "fact". I cannot help feeling, however, that not all facts are equally worth knowing. I thought it was the task of the natural sciences to discover the facts of nature, not to create them.

Y:

You are just an obscurantist.

O:

I have often been accused of spreading darkness. And I cannot

deny that the dazzling light being thrown on a few spots to the exclusion of the rest has distorted all the proportions of our science. How many frantic about-faces someone of my age has had to live through! And the yelping is not at an end; in fact, it is getting worse. Miss Molecule of 1962 remains to be crowned; and I fear—and you hope—there will be many more. The next candidate will no doubt be messenger RNA. And some time later it will be the operon buffo: “Figaro qui, Figaro qua!” But to continue with what I was saying before. You may, of course, ascribe to my advancing years the difficulties I experience in understanding what is claimed to go on. Even if we accept the existence of a short-lived species of RNA that is made under the direct control of DNA and by mirroring the composition of the latter transfers the information from DNA to the proteins and if we add the relatively small amount of so-called soluble RNA, we are left with the bulk, almost 85 per cent, of the cellular RNA for which neither function nor mode of formation is apparent. You will, of course, with your usual originality tell me that Rome was not built in one day, to which I shall reply that less credit should be given to those who fail to solve great problems than to those that succeed in solving small ones. There is nothing easier than to fall off the Mount Everest.

Y:

But you cannot deny the existence of a rapidly turning-over RNA that does mirror the composition of DNA, in base-pairing, etc.?

O:

Most of the analyses that I have seen are far from convincing. A lot of good will or, better, lack of experience is necessary for such enthusiastically sweeping claims. But never mind, many people have made a good living in science by selling the emperor's new clothes. *De nihilo nihil* does not hold for molecular biology. Only where there is nothing, all is possible. If you consider a cell, and the amount of packing or compression it must require, the traffic problems become so enormous as to demand the for-

mulation of a new dimension for which we lack even the slightest intimation; instead of which we are being fed terminological cant. DNA slowly reproducing itself and at the same time rushing to make hundreds of different messengers; all these messengers milling about madly in search of ribosomal easy chairs on which to procreate and die; proteins peeling off and proceeding to their respective posts; unborn lipids and polysaccharides crying for non-existing templates: a molecular Walpurgis Night, a Federation Meeting of the universe, only even less comfortable than in Atlantic City. It would all be very funny if it did not corrupt our youth. Who, when he knew it when it was *so* small, would have thought of DNA as the demiurgos of a Manichaeian world?

Y:

You are using too many long words to say nothing. The fact that you knew DNA when it had a molecular weight of 800,000, whereas now it is up to 160 million, is of no significance, except to show that you did not know how to invest wisely.

O:

Of course, this is not at all what I meant. But even if we reconcile ourselves to this bizarre situation and accept the notion that everything that happens in a cell is under the ultimate control of DNA, we encounter other difficulties. Many viruses are essentially ribonucleoproteins; and in several cases their RNA itself has been shown to be infective. You mentioned it yourself before. This RNA is presumably able to replicate itself; but where is the specific DNA that presides over this replication?

Y:

I can give you two different answers. I could assume that there is something special about viral RNA that enables it to act as its own template, independently of any control exercised by DNA.

O:

But this is a self-strangling argument. Shouldn't in this case viral

RNA show complete base-pairing? Besides, from the chemical point of view, there does not seem to be anything special about the RNA of viruses. Incidentally, since in this reasoning all roles are reversed, I expect to hear almost any day about a "messenger DNA".

Y:

I have an alternative answer. Let us assume that all RNA, viral as well as ribosomal, transfer or messenger, is made under the control of DNA, but only in the case of messenger are both DNA strands involved; the other RNA species reflect the composition of a single strand of DNA. By now you must credit me with enough intelligence to propose hundred plausible schemes to explain this.

O:

I would not go so far; but for n problems you will always have $n + 1$ mutually exclusive, but equally irrefutable, explanations.

Y:

Never mind. So far as virus RNA is concerned, a second assumption must, of course, be made, namely, that for a cell to be able to support a certain virus it must contain, as part of its genome, a stretch of DNA that is normally non-operative, but becomes active under the influence of the invading viral RNA molecule; and it is this DNA portion that is reflected by the infective RNA.

O:

I feel like Peer Gynt in the cave of the trolls, refusing to undergo a simple eye operation that would permit him to see as straight what looks crooked to him. You, too, will undoubtedly grow into one of our synthetic geniuses—these fake celebrities, glued together with the spittle of Madison Avenue—of whom there are so many that they clutter up the place. And yet, the brain power of all these cytopractors combined would not fill the inkwell of

Pascal. To them, everything is so simple. The hypnotic effect of repeating nonsensical statements over and over has produced a general trance that is mistaken for a view of nature. Altogether, I was taught that it is the task of the natural sciences to understand, not to outwit, nature. I am often told that this or that is an "educated guess"—a truly nasty expression. Much would be gained if the guessers were educated instead. Some of the discussions on microbial heredity and chemical genetics that I have heard sounded like a bunch of midwives deliberating on the immaculate conception. And what can be done to stem the ever-growing avalanche of rubbish being published? I can think of only one way: to publish all papers anonymously, without authors' names.

Y:

By the way, grapevine has told me that the code has been broken.

O:

I hope, someone keeps the pieces; one may need them again.

Y:

How can you be facetious before such an achievement? Don't you see that we have entered a new era?

O:

Thus, slowly, slowly, step by step, scream by scream, drum roll by drum roll, gold medal by gold medal, do you expect to reconstruct the fingerprints of God! But what good will it do you? You could not read them, you could not classify them. All the gimmicks in the world—to use the terminology with which you are familiar—will not help you.

Y:

You talk as if we were still in the early middle ages.

O:

Maybe the natural sciences will always be in the early middle

ages. There goes a deep crack through our porcelain world; and even theoretical physics, perhaps the most highly developed science, feels, I am told, much malaise and intellectual discomfort. Could it be that molecular biology is the last refuge of the scientific optimist?

Y:

Well, we have much to be optimistic about. This has been a marvelous period for the biological sciences, a true renaissance. We have learned more about life and heredity in the last five years than in the preceding fifty; and for this reason we can afford to disregard most of the older literature. Even you, who has difficulty in reconciling yourself to all our discoveries, will not be able to deny the tremendous upsurge.

O:

There is so much to get reconciled to. In fact, that such very bad times as ours have given rise to so much good science, does this not speak against science?

Y:

Not at all. You seem to have the romantically foolish idea that only a good man can be a good scientist.

O:

It is always dangerous to use the argument *ad hominem*, and you should not judge from yourself. But is it not a desperate situation when an old proverb must be reversed to read: Wherever the fish stinks there is its head? It is getting late, though, and I had not quite finished with what I was saying before. Even if the correct code is found and the flow of so-called information takes place as postulated by the admirers of biological automation, very little of what occurs in a living cell is really clarified. What determines the specific character of a cell, which is perpetuated in a hereditary fashion, is constituted of a very large number of different compounds, many of them situated specifically within the cell; and these substances, once we break the cell and isolate

and separate its constituents, will be recognized as very many different species of molecules, such as proteins, lipids, polysaccharides, nucleic acids, etc. Most of those—and not only the first and last—have a highly specific composition and structure; but how they exist, interact and are held in definite places in the functioning cell is completely obscure. I am certainly no vitalist in the sense in which this is usually meant; but I cannot stomach people who claim that they have understood and explained “Hamlet” by telling me how often the word “and” appears in the first act. And I do object to the tremendous noise that is being made about trivial and often meaningless observations. In contrast to previous times, I believe, many of our reputedly great discoveries are entirely unearned. Also, the morass of alleged facts in which we are suffocating has brought it about that those who *may* speculate *cannot* do so any longer. Look at the enormous variety in the shapes of organisms, organs, even cellular components—where is the biochemistry of specific shape? Where is the biochemistry of cell differentiation? Is there a separate nucleotide code for your fingerprints which are different from mine? Is it a pairing error in position 79 which has produced the visions of Blake? It is above all against this shabby mechanization of our scientific imagination, which kills all ability to notice the unforeseen, that I protest, against this mat finish over a chaos of unrecognized ignorance, this butcher-like brutality with things that cry for gentle caution. Our young people are being brought up in the belief that “they never had it so good”. They—and especially the best ones—are being condemned to a future of disillusion and discouragement.

Y:

That’s what you say. I am not at all discouraged; quite the contrary. I only have to think of the unprecedented possibilities that are opening before us. When we know the Universal Code, we shall soon learn how to interfere with certain nucleotide sequences in DNA, how to change them specifically and thereby to produce desirable genetic changes. Artificial insemination with the stored

sperm of dead geniuses has already been proposed by eminent authorities. Two little Einsteins in every middle class household, what a vista!

O:

But can the centipede survive under a duodecimal system? The more you talk about breeding geniuses the less likely you are to get them.

Y:

Again one of your unscientifically mystical remarks. I only mentioned this proposal as a first modest feeler into the great future. Later on, we shall be able to look up the nucleotide sequence of every DNA; and each purine and pyrimidine will have a number; and we shall know of each what happens when we change it. And boy, will we change it!

O:

And then you will get the real "human engineering". Once you can alter the chromosomes at will, you will be able to tailor the Average Consumer, the predictable user of a given soap, the reliable imbiber of a certain poison gas. You will have given humanity a present compared with which the Hiroshima bomb was a friendly easter egg. You will indeed have touched the ecology of death. I shudder to think in whose image this new man will be made.

Y:

Well, you may not be around to see it. Anyway, it was nice to have talked with you. I must still run up to the lab to turn off the Spinco.

O:

Yes, the evening has come. I shall go home.

They leave, in opposite directions.

Index

- Acid degradation of DNA, lability of purine glycosides to acid, differential distribution analysis, 123, 124
- Adenine and guanine, content in animal DNA's and in animal RNA's, 6
- Analogues, selection of fraudulent — in formation of DNA, 152, 153
—, — of natural — in formation of DNA, 150–152
- Apurinic acid from calf thymus DNA, composition and properties (table), 33, 34
—, definition, 33
— in sequence analysis, 33, 91, 123, 133–134, 166
- Arbacia lixula*, pyrimidine 3',5'-diphosphates from DNA, differential distribution analysis, 92, 122, 138
- Arrhythmic character of nucleotide sequence in DNA, 91
- AT type DNA, 28, 51, 66, 86, 128, 141, 165, 170, 172
- Autarkic entities (cells and cell communities), and viruses and phages, requirements in —, 71, 101, 103
- Avian tubercle bacilli, composition of DNA (tables), 13, 14, 87
—, ratios in DNA, 30, 50, 66, 87
- 8-Azaguanine, incorporation into DNA, 152, 155
- Azotobacter vinelandii*, invariability of total RNA, 105
—, nucleotide ratios in ribonucleo-proteins, 105, 119
— RNA, 94, 105, 119
- “Backbone” of nucleic acids, 90
- Bacterial transformation by a specific DNA, 71, 121
- Base-pairing regularities in DNA, 13 and *passim*
- Biochemical evolution, 78
— and formation of macromolecules, 78
- Biochemical space, 115–117
- Biochemistry, genetics, and nucleic acids, 163, 164
- Biological information, 73, 95, 113–115, 163, 184
—, chemically conveyed, 102–105
—, DNA as repository of, 104, 165
—, flow of —, 103, 113–115
—, linguistic analogies, 104, 113, 130, 165
— and nucleic acids, 73, 74, 100–107
— and regularities in nucleic acids, 95, 121
—, sequential specificity in proteins, nucleic acids, polysaccharides and preservation and transfer of information, 103
—, triangular relationship between DNA, RNA and proteins in preservation and transfer of —, 103

- Biological polymer, what determines the size of a — ?, 127
- Biological role of nucleic acids, 26, 70–72
- —, participation of proteins in, 26, 73
- 5-Bromouracil, incorporation into DNA, biological consequences, 143–146
- , — of *E. coli* thymine auxotroph, 152–154
- Bunched and solitary nucleotides (pyrimidine), 123, 148, 166
- —, schematic representation of a fragment of a double strand of DNA, 90, 132
- Calf liver, composition of RNA's (table), 17
- Calf thymus DNA, 12, 138
- , apurinic acid from —, 34
- , enzymic degradation, 19, 31, 33
- , fractional dissociation from histone and globin complexes, 45–50
- , method of isolation, 129
- , 5-methylcytosine distribution in, 56, 151
- Cell constituents, spatial conditions under which the cell manufactures its constituents, 116
- Cell and machine, 174, 175
- Cellular specificity, chemical basis, 102–104, 112, 113
- , directive influence of proteins and nucleic acids in maintenance of —, 104
- and heredity, 111
- Chemical concept of a molecule and molecular biology, 182
- Chicken erythrocyte DNA, 30–32
- 5-Chlorouracil, incorporation into DNA, 152, 154
- Citrate ions, use in isolation of nucleic acids for inhibition of deoxyribonucleases, 6
- Clusters of nucleotides in DNA, 19
- —, analysis of sequence, longer all-thymine than all-cytosine units in many DNA's, 149
- Code-script of biological high polymers, 117–124
- Coding of information and order of alignment of the nucleotide monomers in a nucleic acid, 124
- for phenylalanine, 169
- units for amino acids, critical remarks, 168–172
- Comparative biochemistry, 112
- Compositional regularities in nucleic acids 13, 28, 30, 51–52, 54, 66, 72, 86–88, 93–96, 106, 120, 127, 128, 131, 155, 164, 170
- Conjugated proteins, 78–81
- Conjugation in the formation of macromolecules, role in biochemical evolution, 78
- Cybernetics, 113
- Cytidine, isolation and estimation, 9
- Cytochrome *b₂*, polydeoxyribonucleotide component of —, 128
- Cytoplasm of rat liver and kidney, nucleotide ratios in ribonucleoproteins, 94, 119
- Cytosine, decomposition to uracil by heating with strong acid, 10
- Cytostatic and cytotoxic chemicals which are incorporated into nucleic acids, 71
- Deoxycytidine and thymidine 3',5'-diphosphates released from DNA of various organisms (table), 92, 136, 138, 141, 142
- Deoxymethylcytidine 3',5'-diphosphate, 140
- Deoxypentose nucleic acid(s), *see also* deoxyribose nucleic acid(s), deoxyribonucleic acid(s), DNA
- , sugar in —, 18, 53, 65, 84
- Deoxypentose nucleoprotein(s) of calf thymus and wheat germ, course of dissociation at rising NaCl concentrations, 46, 65

Deoxypentose nucleoprotein(s) (*continuation*)

- , classification according to the protein moiety, 41
- , dissociation and cleavage, behavior towards electrolytes and fractionation of DNA, 45, 57, 58, 65
- , fibrous structure, 40
- , isolation by extraction with solutions of low ionic strength, 42, 65, 89
- , — with strong salt solutions, 42, 65, 89
- and their prosthetic groups, 39–60

Deoxyribonuclease from pancreas, action on calf thymus DNA, formation of non-dialyzable core, 18, 19, 28, 31, 33

- , stepwise degradation of DNA, 19, 91, 133
- of yeast, inhibitor specificity, 19, 20

Deoxyribonucleohistone, isolation, 65

Deoxyribonucleoprotamine, isolation, 65

Deoxyribose nucleic acid(s) (deoxyribonucleic acids, DNA, deoxypentose nucleic acids), *see also* nucleic acid(s)

- , absence of organ specificity in composition of DNA's of the same species, 13, 14
- , — of periodicity in structure, 33, 51
- , arrhythmic character of nucleotide sequence, 91
- , AT type and GC type, 28, 51, 66, 86, 128, 141, 165, 170, 172
- , balance between amino groups and enolic hydroxyls in various DNA preparations, 32, 52
- , balance of 6-amino groups and of 6-keto groups, 32, 52
- of calf thymus, degradation by deoxyribonuclease (table), 19

Deoxyribose nucleic acid(s) (*continuation*)

- , common features, 27
- , composition, 11–13
- , — of DNA fractions obtained by fractional extractions of wheat germ nucleoprotein gels (table), 68
- , — in human tissues (table), 12
- , — in normal human tissue and in cancer tissue, 28
- , — and regularities, 86–88
- , — in several organs of the ox (table), 12
- , — of several total — (table), 87
- , — in yeast and avian tubercle bacilli (table), 13
- content per diploid and haploid nucleus, 27, 71
- and copying mechanism, 73
- , differential analysis of pyrimidine distribution in three analytically indistinguishable — (table), 122
- , dissymmetry ratios, 86
- , diversity, 32, 57
- , equality of sum of purine nucleotides and sum of pyrimidine nucleotides, 52, 66
- , evidence for the existence of a very large number of — in nature, 50, 51
- , examples of pleromeric relationships (table), 154, 156
- , fractionation, 57–59, 88–89
- , — through their artificial complexes with histone, globin or polylysine, 57
- , — on histone-kieselguhr column, 57
- , — at the nucleoprotein stage, 66–68
- , genes and specificity of — in chromosomes, 31
- , identity of composition in different tissues of the same species, 27, 28

- Deoxyribose nucleic acid(s) (*continuation*)
- , induction of transformation of bacterial types, 26
 - , inhibition of degradation by deoxyribonuclease with citrate ions, 6
 - , involvement in the genetic apparatus of the cell, 71
 - , isolation of nucleoproteins as first step in isolation of —, 50
 - , location in cells, 25, 26
 - , maintenance of sequence integrity, 129
 - , molar ratios in — of different origin (table), 30
 - , — of adenine to thymine and of guanine to cytosine (+ methylcytosine), 13, 51, 52
 - , molecular structure, discussion of Watson-Crick model, 53-56
 - , molecular weight, 5, 88, 128, 165
 - , position of the phosphate bridges in calf thymus —, 65
 - , presence outside the cell nucleus, 25
 - , purines and pyrimidines, 65, 66
 - , regularities in composition, *see* base-pairing regularities
 - , schematic representation of segment, 90, 132
 - , selection of fraudulent and of natural analogues in formation, 150-153
 - , sequence analysis, *see* sequence analysis of DNA
 - , species differences, 13, 14, 86-88
 - , species specificity and organ specificity, 13, 66
 - , specimens, uniformity of absorption spectrum in the ultra-violet, 32
 - , structure and sequence of nucleotides, feasibility of sequence analysis, 90
- Deoxyribose nucleic acid(s) (*continuation*)
- of thymus, spleen, liver, tubercle bacilli, human sperm; isolation, 7
- Deoxyribothymidine, separation from thymidine, 9
- Depolymerizing enzymes of nucleic acids, 18-20
- Differential distribution analysis of DNA, characteristics of nucleotide sequence, 122-124, 134-148
- —, mechanism, 134-137
- Differential pyrimidine distribution analysis, principal conclusions, 166, 167
- Dinucleoside triphosphates comprising 5-methylcytidine and either cytidine or thymidine, 140
- Dissymmetry ratio of nucleic acids, definition, 66, 86
- Diversity and identity in high-molecular cell constituents, 2-5
- of nucleic acids, chemical diversity in DNA of various species and in total cell DNA, 50, 57, 66, 88, 105-106
 - —, — in RNA of various species and in RNA's of the same cell, 105-106
 - — isolated from a given cell, 66
- Division of labor in the life of the cell, 185
- DNA, *see also* deoxyribose nucleic acid(s) (deoxyribonucleic acid), deoxypentose nucleic acid(s)
- , is there evidence for the existence of DNA chains for biological information in higher forms of life?, 163
- DNA and RNA, critical remarks about present conceptions of functions of —, 185-195
- , remark about the terms, 84
- Dynamic state of the body constituents, 183

- Echinocardium cordatum*, ratio of adenine + thymine to cytosine + guanine in DNA, 50
- End groups in undegraded DNA, absence of — ?, 53, 129
- Escherichia coli*, DNA produced by thymine auxotrophs in the presence of bromouracil, 143–147
- , equimolarity of the bases in DNA, 51, 86
- , invariability of total RNA, 105
- , K-12 strain DNA, 30, 31, 86, 87
- , longer all-thymine than all-cytosine stretches in DNA, 149
- , molar ratios of bases present in DNA, 51, 86, 87
- , nucleotide ratios in nucleoprotein, 94, 119
- , — in RNA, 69, 119
- , pleromeric relationships in DNA, 154
- protoplasts, method of isolation of DNA, 130
- , transformation by DNA, 26
- 5-Fluorouracil, incorporation into DNA, 152, 156
- Fractionation of deoxypentose nucleic acids at the nucleoprotein stage, 57, 58, 66–69, 121, 123
- GC type DNA, 28, 31, 51, 66, 86, 128, 141, 142, 164, 165, 170
- Genes and specificity of DNA in chromosomes, 32
- Genetic apparatus of the cell and DNA, 71
- Genetic determinants and nucleic acid chains, many determinants in one DNA chain or one DNA chain for each determinant, 121
- Genetics, biochemistry, and nucleic acids, 163, 164
- Harmony of cellular life, 4
- Hemophilus influenzae*, transformation by DNA, 26
- Hemophilus influenzae* (continuation)
- , type *b* and *c*, DNA, 26
- , type *c*, DNA, 30, 31
- Heredity, definition, 111
- , first steps towards a chemistry of —, 109–124
- Heterogeneity of nucleic acids as regards composition and sequence characteristics, 128
- Hierarchy, existence of — in the cell, 1, 63, 100, 101, 117, 185
- High-molecular cell constituents, identity and diversity, 2–5
- Homogeneity of nucleic acids in the cell, 127–128
- Homotopes and pleromers in DNA's, 157
- in polymer constitution, definition, 131
- Homotopy and pleromerism and the neighbor problem in formation of nucleic acids, 153–158
- of thymine and 5-methylcytosine in rye germ DNA, 155
- Human tissues, composition of DNA from various — (table), 15
- , differential distribution analysis of DNA of spleen (table), 138
- , isolation of sperm DNA, 7
- , molar ratios of nucleotides in DNA of spleen, liver and other tissues, 12, 14, 30, 87, 92
- , pyrimidine 3',5'-diphosphates from DNA, differential analysis, pyrimidine distribution, 92, 122, 138
- Hydrolysis of the nucleic acids, 10–11
- 5-Hydroxymethylcytosine, 86
- in bacteriophage nucleic acids, 52
- in DNA of *E. coli* phages, 66
- Identification of the macromolecular compounds of cellular origin, 3, 4, 104

- Identity and diversity in high-molecular cell constituents, 2-5
 — — —, is there a critical range of molecular weights as for ability of cells to synthesize completely identical substances?, 5
 Information, *see* biological information
 Information theory, 113
 Invariability of nucleic acids, 104-105
 — within the species, 120
 5-Iodouracil, incorporation into DNA, 152, 154
 Isolation of nucleic acids, exclusion of degradation, 6

 Lipoproteins, nature of bonds between lipid and protein, 80
 —, specific patterns, 80, 113
 Liver, nucleic acids, 11, 12, 14, 15, 17, 93

 Macromolecular structure, the problem of the — of the native state of DNA, 129, 130
 Magnesium ions, effect on fractional dissociation of artificial complexes of histone with calf thymus DNA, 48
 — requirement by deoxyribonucleases, 6
 Mechanomorphic concept in biochemical conceptions of cell life, 103
 — of living nature, 111, 175
 6-*N*-Methyladenine, 86, 153
 Methylcytosine-cytosine units in rye germ DNA, 157
 5-Methylcytosine, amount in DNA of wheat and rye germ (table), 139
 — distribution in fractions of calf thymus DNA (table), 56
 — in DNA of *E. coli* phages, 66
 —, frequency of — and cytosine as solitary and bunched nucleotides in DNA, 152
 5-Methylcytosine, frequency of — (*continuation*)
 —, — and thymine as solitary nucleotides in DNA, 152
 —, high proportion linked to guanine in a dinucleotide as compared with cytosine in various DNA's, 56, 157
 —, occurrence, 86
 —, percentage contribution to aminopyrimidine content in various fractions of DNA, 151
 — in structure analysis of nucleic acids, 139-141
 — in wheat germ DNA, 52, 55, 139
 Minimum of signaling potential in biological information, 117
 — of viable organization, 117
 Models and model experiments, reflections on —, 53, 77, 83, 95, 174, 186, 187
 Molecular weights of DNA, 5, 88, 128, 165
 Morphogenesis, role of RNA, 22, 71
 Mutagenic action spectrum of ultraviolet radiation and absorption spectrum of nucleic acids, 71
 Mutagenic effects of chemicals which react with nucleic acids, 71
Mycobacterium phlei, nucleotide ratios in RNA, 69, 94, 119
Mycobacterium tuberculosis (bovine), nucleotide ratios in RNA, 119

 Native DNA, preservation of native state, 129
 Native state of nucleoproteins, criteria, 42, 43
 Neighbor principle, high tendency of methylcytosine to be next to guanine in DNA's, 157
 — in nucleotide sequence in nucleic acids, 153-158
 — in soluble RNA of liver cytoplasm, 157
 — in structure of calf thymus DNA, 72

- Nomenclature, remarks on — in sequence analysis of nucleic acids, 131–132
- Nuclei of rat liver, nucleotide ratios in ribonucleoprotein, 119
- Nucleic acid(s), *see also* deoxyribose nucleic acid(s) and pentose nucleic acid(s)
- , absence of repeating polynucleotide units, 51, 72
 - , biological role of — and of nucleoproteins, 26, 73
 - , brief listing of relevant information on the biological roles of the —, 71–72
 - , as carriers of biological information, 73, 100–107
 - , depolymerizing enzymes, 18–20
 - , diversity, nucleotide sequence and periodicity in — chain, 31, 32, 33, 105–106, 119
 - , exclusion of degradation by enzymes, acid or alkali, in isolation, 6
 - , invariability, 104–105, 120
 - , isolation, 6
 - , meaning of regularities in composition, 93–96
 - , methods for analysis of nucleotide composition, 6–11
 - , methods of hydrolysis, 10–11
 - , principal constituents, 85, 86
 - , reaction with chemicals and mutagenic effects, 71
 - , reflections on possible role of DNA in transmission of inherited properties and of pentose — in differentiation, 21, 22
 - , remarks on functions, 70–74
 - , remarks on nomenclature in sequence analysis, 131–132
 - , role of proteins in function of —, 73, 74
 - , separation at the nucleoprotein stage, 68
 - , some immediate problems in the chemistry of —, 167, 168
- Nucleic acid(s) (*continuation*)
- , structure and function of — as cell constituents, 25–36, 70–74
 - , — and sequence of nucleotides, 90
 - , sugar components, 18, 53, 65, 84
 - , ultraviolet absorption spectrum and mutagenic action spectrum of ultraviolet radiation, 71
 - , urgency of sequence analysis, 73
- Nucleohistone(s), 41, 64, 79
- , bonds between nucleic acid and histone, 41
 - of calf thymus, isolation, 42
 - , fractional dissociation of calf thymus — and of wheat germ nucleoprotein, 46, 89
 - , fractionation of nucleic acids in — by exposure to increasing NaCl concentrations, 44, 45, 65
 - , nucleic acid content in —, 43
 - , properties of calf thymus — solutions at different NaCl concentrations (table), 44
 - , solubility, 43
- Nucleoprotamines, 41
- , bonds between nucleic acid and protamine, 79, 80
 - , dissociation in strong salt solutions, 43, 45
 - , phosphorus to arginine ratio, 43
 - from sea-urchin sperm, isolation, 42
- Nucleoproteins, *see also* ribonucleoprotein(s) and deoxypentose nucleoprotein(s)
- , absorption spectrum in the ultraviolet, 43
 - , artificially prepared complexes of nucleic acids with positively charged histone, globin and synthetic polylysins and polyvinylamine, 45–49, 57, 58
 - from avian tubercle bacilli, isolation, 42
 - , bonds between nucleic acid and protein, 64, 79, 80

Nucleoproteins (*continuation*)

- , deoxypentose — and their prosthetic groups, 39–59
- , distinction between pentose — and deoxypentose —, 64, 79, 80
- as fibrous proteins, 40
- , fractional dissociation of artificial complexes, 46, 47
- , function of the nucleic acid and of the protein part in —, 107
- , native state, 42, 43
- , nucleic acid content, 43
- , remarks on —, 96
- and RNA's, 92–96

Nucleotide arrangement in different DNA's of animal origin (table), 137, 138

- in the DNA of *E. coli* grown in the absence and presence of 5-bromouracil, 142–146
- in a microbial DNA of the GC type (table), 140–142
- in the total deoxyribonucleic acid of rye germ and in a complete series of fractions, 137–140

Nucleotide ratios in ribonucleoproteins, 69, 94, 119, 120

Nucleotide sequence in DNA (*see also* sequence analysis), 126–158

Order of alignment of the nucleotide monomers in a nucleic acid chain and coding of information, 123

Organ specificity of composition of nucleic acids, 12, 66

Ox tissues, composition of DNA of thymus, spleen, liver, 12, 14, 15, 30, 87

—, molar ratios in DNA, 14, 15, 30, 87, 92

—, nucleotide ratios in ribonucleoproteins of liver, 94, 119

—, — in RNA of liver, 17, 69

—, pyrimidine 3',5'-diphosphates from DNA, differential analysis, pyrimidine distribution, 92, 122, 138

Paper chromatography of purines and pyrimidines, 5, 7–9, 149, 180

Paracentrotus lividus, differential distribution analysis of sperm DNA, 138

— DNA, 87

—, eggs and embryos, nucleotide ratios in ribonucleoproteins, 69, 94, 119

Pentose nucleic acid(s) (PNA, ribonucleic acids, RNA)

—, action of ribonuclease, 34, 35

—, composition, 15–17, 93

—, — in animal tissues (table), 17

—, — in several preparations from yeast and liver (table), 93

—, — in yeast (table), 16

—, differences between — of the same organ of different species and of different organs of the same species, 17

—, — in composition between nuclear and cytoplasmic RNA in liver, 25

—, — between RNA's from different locations within the cell, 69–70

— and differentiation, 22

—, equality in the proportion of 6-amino and 6-keto compounds and system for transfer of information, 52, 66, 74, 93, 106

—, frequency distribution of various nucleotide ratios in total — of 104 samples, 94

—, guanine content of RNA of pancreas, 17

—, isolation from animal tissues, 7

—, location in cells, differences between nuclear and cytoplasmic —, 25

—, methods of analysis, 15, 16

—, — of hydrolysis, 15, 16

—, molecular weight, 4

— and nucleoproteins, 92

—, nucleotide ratios in various organisms and organs (table), 69

—, preparation, 7

- Pentose nucleic acids (*continuation*)
—, regularities in composition, 52, 66, 74, 93, 106
—, role in protein synthesis and in morphogenesis, 71
—, soluble, for transfer of amino acids in liver cytoplasm, 157
—, species and organ specificity, 17
—, sugar component in —, 18, 84
Pentose nucleoproteins, fibrous, 40
—, links between nucleic acid and protein, 64, 80
—, occurrence, 64, 65
Periodicity, absence in nucleic acids, 31, 33, 91
Phage T_{2r}⁺ DNA, pleromeric relationships, 154
Phosphoric acid bridges in DNA, 53, 65, 89
Pig tissues, action of ribonuclease on PNA, 34
—, fractional dissociation of liver DNA from histone and globin complexes, 46
—, RNA of liver, pancreas, 17
Pleromer, definition of the term — in polymer constitution, 131
Pleromeric relationships in DNA's (table), 154
— in RNA's, 155, 156
Pleromerism and homotopy and the neighbor problem in nucleic acids, 153–158
Pleromers and homotopes in DNA's, 157
Pneumococcus, transformation by DNA, 1, 26, 71, 121, 181
Precursors, movement of — into various polymers in cellular elements, 115, 116
Prosthetic groups in conjugated proteins, 81
Protein synthesis, role of nucleic acids, 71, 95, 96
Purines, location on paper chromatograms in ultraviolet light, 9
Purines (*continuation*)
—, molar proportions in DNA's from different organs of one species (table), 15
—, — — from different species (table), 14
—, release from nucleic acids by mild acid hydrolysis, 9, 10
— satellites in DNA's, 52
—, schematic representation of the position on paper chromatogram, 8
—, separation and estimation, solvent systems, 7–9
—, spectrophotometric determination, 9
Pyrimidine(s) distribution of the total DNA of rye germ and wheat germ and its fractions, 139–141
—, location on paper chromatograms in ultraviolet light, 9
—, molar proportions in DNA's from different organs of one species (table), 15
—, — — from different species (table), 14
— nucleoside diphosphates, formation by acid treatment of DNA, use for study of the nucleotide sequence, 91, 92, 123, 134, 135
— nucleotides, all-thymine and all-cytosine units in DNA, 149
—, liberation of the pyrimidines, 10, 11
—, solitary and bunched — in DNA, 123, 148, 166
—, occurrence in DNA as oligonucleotide tracts containing three or more pyrimidines in a row, 148
—, release from nucleotides by hydrolysis with concentrated formic acid, 11
—, resistance to treatment with strong acid (table), 10
—, schematic representation of the position on paper chromatogram, 8
—, separation and estimation, 7–9

Pyrimidine(s) (*continuation*)

—, spectrophotometric determination, 9

Rat tissues, nucleotide ratios in ribonucleoproteins of liver and kidney and cytoplasm of liver and kidney, 69, 94, 119

—, pleromeric relationships in liver DNA, 156

—, uptake *in vivo* of radioactive leucine by liver cytoplasm, 116

Reflections on present trends in biological sciences, 100–102

Regularities in composition of nucleic acids, meaning, 93–96

— — of RNA's and DNA's, 13, 28, 30, 51–52, 54, 66, 72, 86–88, 93–96, 106, 120, 127, 128, 131, 155, 164, 170

Replacement, meaning of the term, — of natural purine or pyrimidine in nucleic acids by analogues, 150

Ribonuclease, action on pentose nucleic acid of yeast and pig liver, composition of dialyzable nucleotides and core (table), 34

— from pancreas, 18, 34, 35

Ribonucleic acid, *see* pentose nucleic acids, RNA, PNA

Ribonucleoprotein(s), links between nucleic acid and protein, 64, 80

— of the ribosomes, 165

Ribonucleotides, methods for separation and quantitative determination in an aqueous ammonium isobutyrate-isobutyric acid system, 9

Ribosomes, 165

RNA (*see also* pentose nucleic acids, ribonucleic acids, PNA)

— and DNA, remark about the terms —, 84

Rye germ DNA, comparison of the pyrimidine distribution in total DNA with that in DNA fractions, 139–141

Rye germ DNA (*continuation*)

—, pleromeric relationships, 154

Salmon DNA, identity of composition as regards distribution of nitrogenous constituents with sheep DNA, 137

—, molar ratios, 30

Salmon sperm DNA, composition, 29, 30, 87

Sea urchin DNA, dissymmetry ratio, 66

— sperm DNA, 42, 66, 138

Sequence analysis of DNA, 31, 33, 90, 91

— of nucleic acids, lines of attack, 91, 92

— —, possibilities of cluster analysis, 148–150

—, remarks on the conceptual basis of —, 127–128

Sequence characteristics common to all DNA's, 147–148

— of DNA, occurrence of at least 70% of the pyrimidines in animal and plant DNA as oligonucleotide tracts containing three or more pyrimidines in a row, 148

— of nucleic acids, 124

Sequence investigation, apurinic acid, 133, 134

—, early attempts, 132

— of nucleic acids through differential distribution analysis, *see* differential distribution analysis

—, the problem of statistical analysis, 130–131

—, stepwise degradation of DNA by deoxyribonuclease, 133

Sequence of nucleotides in DNA, the problem of non-randomness, 147, 148

Sequence studies of DNA, difference in stability to acid of purine and pyrimidine nucleosides in DNA, 135, 166

- Sequential specificity of high polymers as mark of the specificity of the cell, 165, 166
- Serratia marcescens* DNA, 30
- Sheep DNA, 87, 137
- Sheep liver, composition of RNA's, 17
- Solitary and bunched nucleotides, in DNA, 166
- —, schematic representation of a fragment of a double strand of DNA, 132
- Soluble RNA's of liver cytoplasm for transfer of amino acids, 157
- Solvent systems for separation of purines and pyrimidines, 7-8
- Species diversity and invariability of composition within the same species of highly polymerized cell constituents, 120
- Species specificity of composition of DNA, 13, 15, 66, 120, 164
- — of RNA, 120
- Specific arrangement of the constituents of cell-specific polymers, role of nucleic acids, 118
- Specificity, direct gene-controlled and reflected — of highly polymerized cell constituents, 120
- Spermatozoa, ratios of adenine to guanine and of thymine to cytosine in deoxypentose nucleic acids of human —, 12, 13, 14, 15
- Sub-units in DNA, absence of —, 133
- Sugar components of nucleic acids, 18, 53, 65, 84
- Templates, 2, 74, 81, 95, 104, 111, 116, 127, 184, 194
- in time for manufacturing of cell constituents, 81, 116
- Terminal phosphate group in DNA, 53
- Tetranucleotide hypothesis, 2, 3, 13, 28, 51, 85, 127
- Thymic acid, 33
- Thymonucleic acid, 3
- Thymus, nucleic acids of —, 3, 12, 14, 15
- Transformation, bacterial, by DNA, 26, 71, 121, 181
- Tubercle bacilli, distribution analysis of DNA of avirulent variant (BCG) of bovine —, 141, 142
- , isolation of DNA *via* the nucleoprotein, 7
- Ultraviolet absorption spectra of DNA specimens, 32
- of nucleoproteins, 43
- Ultraviolet radiation, mutagenic action spectrum and absorption spectrum of nucleic acids, 71
- Unity of biochemistry, the concept of —, 179
- Unity of nature, the concept of —, 179
- Uracil, instead of thymine in RNA's, 11
- Uridine, isolation and estimation, 9
- Uridylic acid, refractoriness to hydrolysis by acids, 16
- Vitalism, 110
- Wheat germ DNA, 29, 30, 31, 33, 46, 68, 87, 139, 154
- —, composition of intact preparations and of enzymically produced cores, 28, 29
- , nucleotide ratios in pentose nucleoproteins, 94, 119
- RNA, 69
- Yeast DNA, composition (table), 14, 30
- RNA, action of ribonuclease, 34
- —, composition, nucleotide ratios, 16, 69, 93, 94, 119
- X-ray investigations of DNA, 53

PRINTED IN THE NETHERLANDS
BY VONK & CO'S DRUKKERIJ, ZEIST