Papers on DNA composition, 1950-'53 [Papers]

Publication/Creation

19 June 1950 - 11 April 1953

Persistent URL

https://wellcomecollection.org/works/sdpbv5te

Wellcome Collection
183 Euston Road
London NW1 2BE UK
T +44 (0)20 7611 8722
E library@wellcomecollection.org
https://wellcomecollection.org

[From THE BIOCHEMICAL JOURNAL, Vol. 55, No. 5, pp. 774–782, 1953]
 PRINTED IN GREAT BRITAIN

The Bases of the Nucleic Acids of some Bacterial and Animal Viruses: the Occurrence of 5-Hydroxymethylcytosine

By G. R. WYATT
Laboratory of Insect Pathology, Sault Ste Marie, Ontario, Canada

and S. S. COHEN
Department of Pediatrics, Children's Hospital of Philadelphia and Department of Physiological Chemistry, University of Pennsylvania, Philadelphia, Pennsylvania

(*Received* 11 *April* 1953)

Recent studies on the multiplication of viruses have directed attention increasingly toward their nucleic acids. Hershey & Chase (1952) have shown that most, if not all, of the sulphur-containing protein of coliphage T2, which appears to be present in the outer shell of the virus, does not enter the infected cell. However, deoxyribonucleic acid (DNA), apparently organized within the virus, is in some way transferred to the host cell, and appears, therefore, to participate more intimately in the transmission of genetic properties. On infection of *Escherichia coli* with bacteriophage T2, T4 or T6, there is immediate cessation of synthesis of ribonucleic acid (RNA) and net synthesis of DNA is detectable in about 10 min. (Cohen, 1947, 1951). A similar apparent redirection of DNA synthesis during virus multiplication is characteristic of certain induced lysogenic systems, but in this case synthesis of RNA

continues unimpeded (Siminoviteh & Rapkine, 1952). Much work has been directed toward tracing the origin of the structural components of phage DNA (e.g. Weed & Cohen, 1951; Putnam, 1952).

It is important, therefore, to know what chemical properties the nucleic acids of bacterial and other viruses may possess. Wide variations have been demonstrated in the composition, with respect to purine and pyrimidine bases, of the RNA's of plant viruses (Markham, 1953) and the DNA's of insect viruses (Wyatt, 1952*b*).

Concerning the DNA of coliphages T2 and T6, conflicting results have appeared. Smith & Wyatt (1951) reported the presence of substantial amounts of cytosine in phage T2; Weed & Cohen (1951) reported isolation of deoxycytidylic acid from T6; Marshak (1951), however, could find only adenine, guanine and thymine in the nucleic acid of T2, and concluded that this virus contained only these three bases. In no case was the total recovery of nitrogenous bases in terms of phosphorus recorded; in the investigation of Weed & Cohen, a total analysis was not the object, and in the other two studies it was assumed that procedures found satisfactory with DNA from other sources would give quantitative results with phage DNA also.

We have re-examined the DNA's of phages T2, T4 and T6, and found that they do contain no cytosine but instead a hitherto unrecognized pyrimidine base, now identified as 5-hydroxymethylcytosine. We report the quantitative purine and pyrimidine composition of the DNA's of these viruses and also qualitative and quantitative results on some other viruses which were examined with the object of determining the distribution of the new base. Preliminary notices from this investigation have already appeared (Wyatt & Cohen, 1952, 1953).

MATERIALS

Phages T2, T4 and T6

The six viruses which provided the basis of these studies were the r and r⁻ strains of coliphages T2, T4 and T6. Their properties and the isolation of many of the preparations used have been described (Cohen & Arbogast, 1950*b*). Various new preparations were isolated from lysates of high titre (5×10^{11}–10^{12} virus particles/ml.) obtained by growing *Esch. coli*, strain B, to 3×10^8 bacteria/ml. and infecting these cells with an average of five virus particles each. The cells grew exponentially to this level when suspended in 350 ml. of medium at 37° in a 2 l. flask rotated in such a fashion that the medium was thinly layered on its inner surface, thereby providing maximal aeration. After infection, rotation was continued until lysis. The glucose and nitrogen content of the basal mineral medium were fortified tenfold.

In addition, two large preparations of phage T6r⁺, totalling 35 g., were generously prepared for us by Dr J. Spizizen of the Research Division of Sharp and Dohme, Inc. Because of difficulties attending the purification of large quantities of virus, these preparations contained a certain proportion of non-virus material. They contained about 20% of phage DNA.

For isolation of DNA the phages were disrupted by urea (3·6 g./10 ml. of virus suspension: Cohen, 1947) and deproteinized in M-NaCl-urea solution by $CHCl_3$:n-octanol (8:1, v/v). The nucleic acids were precipitated with 4 vol. of cold ethanol and washed in 80%, 90% and absolute ethanol and ether. The fibrous solids were dried *in vacuo* over P_2O_5.

Other viruses

Phage T7. Two large preparations containing *Esch. coli* phage T7 were made available to us by Dr J. Spizizen. They contained about 4% of DNA. From one of these, nucleic acid was isolated by the urea technique. A small preparation (5 mg.) of T7 virus was also provided by Dr L. M. Kozloff of the Department of Biochemistry, University of Chicago. This contained about 10% of DNA.

Phage T5. A preparation of this coliphage, containing about 26% of DNA, was the gift of Dr G. Lark, of New York University School of Medicine. From a portion of it, DNA was isolated by the procedure described by Smith & Wyatt (1951), modified by incubating in N-NaOH at room temperature (20°) instead of 37°. NaOH was used in preference to urea in case traces of RNA might be present. We established, using DNA of phage T6, that 5-hydroxymethylcytosine withstands this treatment without loss.

Polyhedral virus. Polyhedral virus from caterpillars of *Colias philodice eurytheme* Bdvl., for which we are indebted to Dr E. A. Steinhaus of the University of California, was isolated by the method of Bergold (1947, 1953).

Vaccinia virus. This material (160 mg., containing about 2·1% of DNA) was prepared for us by Dr A. L. Brown of the Research Division of Sharp & Dohme Inc., using the method of Hoagland, Smadel & Rivers (1940). Treatment of this virus with urea failed to release nucleic acid. DNA was therefore isolated by the procedure of Smith & Wyatt (1951), incubating in N-NaOH at 20°.

Meningo-pneumonitis virus. A preparation of this virus, a member of the psittacosis group, was the gift of Dr M. M. Sigel, of the Research Department, Children's Hospital of Philadelphia. The virus was isolated by centrifugation from chick allantoic fluid, and we were supplied with 17 mg., containing about 2·5% of DNA.

Pyrimidine derivatives

5-Hydroxymethylcytosine and 5-hydroxymethyluracil were synthesized at our suggestion and kindly made available to us by Dr C. S. Miller, of the Research Division of Sharp and Dohme, Inc. The former was prepared* by reduction of ethyl cytosine-5-carboxylate with $LiAlH_4$, and the latter by the method of Litzinger & Johnson (1936).

A specimen of 5-hydroxyuracil, prepared by the method of Davidson & Baudisch (1925), was the gift of Dr G. H. Hitchings. A specimen of 4-methylcytosine, prepared by Dr Hitchings by a method analogous to that used by Hitchings, Elion, Falco & Russell (1949) for cytosine, was obtained through the courtesy of Dr A. Bendich.

* A description of this synthesis is in preparation by Dr Miller and will be submitted to the *Journal of the American Chemical Society*.

EXPERIMENTAL AND RESULTS

Evidence for the presence of an unknown component in phages T2, T4 and T6

Since hydrolysis with perchloric acid (70% at 100° for 1 hr.: Marshak & Vogel, 1951) had proved satisfactory for quantitative liberation of DNA bases from insect viruses, this method was tried first for analysis of phages T4 and T6. The bases were separated by paper chromatography using an isopropanol:water:HCl mixture as the solvent, and eluted and estimated as previously described (Wyatt, 1951*b*, 1952*b*). The principal products were guanine, adenine and thymine, in total amount equivalent to 75–80% of the virus P. The chromatograms also showed weak spots of a substance having an R_P value equal to that of cytosine (0·46–0·50), the yield of which (1–5 moles/100 g. atoms P, using the extinction coefficient of cytosine) varied widely from one experiment to another. On elution, this substance proved to have its ultraviolet-absorption maximum at a slightly longer wavelength than has cytosine, and on being rechromatographed with 86% (v/v) aqueous n-butanol as the solvent its R_P (about 0·07) was much smaller than that of cytosine. The substance was largely unchanged by heating with formic acid at 175° for 2 hr. and so was unlikely to be a nucleoside. In addition, its ultraviolet-absorption spectrum exhibited a shift in alkaline solution similar to those of cytosine and 5-methylcytosine, which the ribosides and deoxyribosides of these bases lack.

It was then found that by using formic acid (88% at 175° for 30 min.) for hydrolysis of the viruses the yield of the unknown was much increased (8–12 moles/100 g. atoms P), the total recovered bases now corresponding to 85–90% of virus P. These observations suggested the presence in phage nucleic acid of a relatively labile pyrimidine base different from any previously reported.

A small amount of the substance, obtained by evaporation of the eluate of a chromatogram spot, was deaminated by treatment with HNO_2 (35 µl. of 2M-$NaNO_2$ and 7 µl. of glacial acetic acid were added and, after standing overnight at 27°, the solution was applied directly to paper for chromatography). The product bore a similar relationship, in spectral and chromatographic properties, to uracil as the parent substance did to cytosine.

The structure of the unknown was suggested to us by study of its spectral and chromatographic properties, as already briefly reported (Wyatt & Cohen, 1952, 1953). Ultraviolet-absorption spectra were read at pH 1, pH 7–8 and pH 13 (Fig. 1, Table 1). From among a number of pyrimidine bases which we examined or whose spectra have been published, only cytosine and 5-methylcytosine exhibited shifts of absorption maximum with change of pH closely paralleling those of the unknown. This suggested that our substance had the polar substituents (2-hydroxy-6-amino-) of cytosine and was additionally substituted in the 5-position. Substitution at C-4 (e.g. 4-methylcytosine) diminishes the extent of the spectral shift in alkali. Addition of an ionizable group at C-5 (e.g. 5-hydroxyuracil) changes the spectrum radically, hence the required substituent should be non-ionizable.

The behaviour of the unknown on paper chromatograms in solvents differing as to pH and to water content, as described in our earlier communications, also suggested ionizable groupings identical with those of cytosine and further indicated that the additional substituent should be hydrophilic in nature. As 5-hydroxycytosine had been eliminated, 5-hydroxymethylcytosine seemed a probable structure.

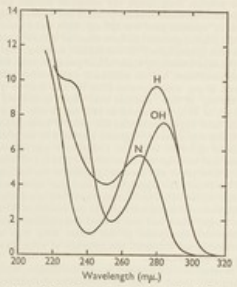


Fig. 1. Ultraviolet-absorption spectra of 5-hydroxymethylcytosine in 0·1N-HCl (H), 0·1N-NaOH (OH) and at pH 7·4 (N). ε is the molecular extinction coefficient. See Table 1.

Table 1. *Ultraviolet-absorption data*

(ε is the molecular extinction coefficient.)

		Maxima		Minima	
Substance	Solvent	Wavelength (mµ.)	ε	Wavelength (mµ.)	ε
5-Hydroxymethylcytosine	0·01M Sodium phosphate buffer (pH 7·4)	269·5	5710	251	4060
	0·1N-HCl	279·5	9700	241·5	1230
	0·1N-NaOH	283·5	7590	254	1890
5-Hydroxymethyluracil*	Buffer (pH 7·4) as above	261	—	231	—
	0·1N-HCl	261	—	231	—
	0·1N-NaOH	285	—	245	—

* Measured from eluates of chromatogram spots; the substance was not isolated in crystalline form.

Isolation and characterization of 5-hydroxymethylcytosine from phage T6

To establish with certainty the identity of the supposed new base it was desirable to isolate in pure form a sufficient quantity for elementary analysis and comparison with synthetic material. For this we were fortunate in having the large phage preparations provided by Dr Spizizen. Since the lability of the new base, unique among natural pyrimidines, was not at first fully realized, yields were much lower than they might have been. Nevertheless, a satisfactory sample was isolated by the following procedure.

From 4·5 g. of the virus concentrate, DNA was prepared by the urea technique referred to above. The product weighed 665 mg. and had N/P = 3·85 (atomic proportions).

For removal of purines (Tamm, Hodes & Chargaff, 1952), 500 mg. of T6 DNA were dissolved in 50 ml. of water and dialysed for 24 hr. at 32° against two changes (each 1000 ml.) of 0·03 N-HCl (pH 1·6), followed by dialysis for 3 days against three changes of distilled water at room temperature. The solution from inside the cellophan dialysis sac was concentrated under reduced pressure and freeze-dried. The yield of 'apurinic acid' was 235 mg., from which some 80% of the original purines had been removed without change in the ratio pyrimidines/P. It was evident, however, that some pyrimidine-containing material had passed through the dialysis membrane, and an additional yield corresponding to about 45 mg. was recovered from the last dialysis water.

220 mg. of this 'apurinic acid' were hydrolysed with 1·5 ml. of 88% formic acid at 165° for 20 min. It is clear from later experiments that the yield of hydroxymethylcytosine would have been greatly increased by use of a larger volume of formic acid. The hydrolysate was spread as a band near one edge of a sheet (18·25 × 22·5 in.) of Whatman no. 3 filter paper and chromatographed in the isopropanol: HCl solvent. The appropriate band was eluted and the eluate concentrated and chromatographed in a similar way with water-saturated *n*-butanol as the solvent on Whatman no. 3 paper which had been washed with distilled water. The ultraviolet absorption of the eluate from this second chromatogram indicated the presence of some 12 mg. of pyrimidine derivative.

Several fractions obtained in a similar manner, containing altogether about 30 mg. of pyrimidine derivative, were combined and treated with 60 mg. of picric acid. The picrate crystallized in needles. After removal of picric acid (by extraction with toluene and ether in the presence of H_2SO_4, the latter being then removed with $Ba(OH)_2$) the solution was concentrated and the free base crystallized in prisms (21·5 mg.). After two recrystallizations from water the product weighed 15·4 mg. Its elementary composition was compatible with the proposed structure. (Found, after drying *in vacuo* at room temperature: C, 40·2; H, 5·35. $C_5H_5O_2N_3$. 1/2 H_2O requires C, 40·0; H, 5·37%. Loss of weight on drying *in vacuo* at 100° over P_2O_5, 4·7% (required: 6·00%). Found, on exhaustively dried material: N (Dumas), 29·3. $C_5H_5O_2N_3$ requires N, 29·8%.)

At approximately the same time as the natural substance was obtained in pure form, synthetic 5-hydroxymethylcytosine was made available to us by Dr Miller. With respect to their ultraviolet-absorption spectra (Table 1, Fig. 1) and to their movement on paper chromatograms in several solvents, the natural and synthetic substances were identical. On heating, both specimens charred slowly above 200°, without melting up to 300°. 5-Hydroxymethyluracil, subsequently prepared by Dr Miller, was similarly indistinguishable from the deamination product of the natural base. We conclude that the base isolated from bacteriophage T6 is 5-hydroxymethylcytosine.

Quantitative hydrolysis of bacteriophage deoxyribonucleic acid

In view of the failure to obtain good yields of 5-hydroxymethylcytosine from DNA by hydrolysis with $HClO_4$, it was of interest to determine the effect of this acid on the pure base. Accordingly, 0·5 mg. of the synthetic base was heated with 0·05 ml. of 72% $HClO_4$ at 100° for 1 hr. Unexpectedly, 97% of the initial 5-hydroxymethylcytosine was recoverable after heating. However, since the nucleoside and nucleotide of this base are not found after hydrolysis of phage DNA with conc. $HClO_4$ at 100°, the low recovery must result from destruction and not from incomplete liberation. This loss may be due either to weakening of the pyrimidine ring by the glycosidic linkage at N·3 or to the effect of other substances in the hydrolysate, such as H_3PO_4. We have not tested these possibilities experimentally.

For quantitative analysis of the DNA of phages T2, T4 and T6, we hydrolysed with formic acid. When a proportion of formic acid to nucleic acid was used similar to that previously used for DNA from other sources, the yield of hydroxymethylcytosine was somewhat variable, and the total recovered bases were equivalent to only some 90% of the DNA P. When pure hydroxymethylcytosine was subjected to these hydrolytic conditions in the presence of thymus DNA, then separated from the other bases by 2-dimensional chromatography and estimated, only some 70–75% of the added amount was recovered (Table 2, expts. 1 and 2). Alteration of time or temperature of

Table 2. *Recovery of 5-hydroxymethylcytosine subjected to hydrolytic conditions in 88% formic acid at 175° for 30 min. in the presence of thymus deoxyribonucleic acid*

Expt. no.	Formic acid (µl./mg. DNA)	5-Hydroxymethylcytosine (moles/g. atom P) Added	Recovered	Recovery (%)
1	80	0·146	0·099	68
	80*	0·146	0·066	45
2	80	0·175	0·134	77
	80*	0·175	0·108	62
3	260	0·215	0·187	87
4	360	0·206	0·195	95
5	600	0·330	0·318	96·5

* Protein (bovine serum albumin, equal in wt. to the DNA) was also added.

hydrolysis did not lead to significant improvement, but recovery of this base was found to depend markedly on the volume of formic acid used. With a sample of 1·5 mg. of phage, or 0·7 mg. of phage DNA, for which 0·05 ml. of 88% formic acid had been used, maximal yields in terms of P are obtained with 0·25–0·5 ml. By using sufficient formic acid, recovery of hydroxymethylcytosine subjected to hydrolysis along with thymus DNA was raised to approximately 96% (Table 2), and the total base recovery from phage DNA became equivalent to 97–99% of total P.

It has also been found important to avoid an excessive volume of air above the formic acid during hydrolysis. This leads to loss of hydroxymethylcytosine, which can, however, be prevented by replacing the air in the tube with nitrogen or formic acid vapour before sealing it off. Evidently the compound is most stable under reducing conditions, and this may account for the efficacy of formic acid in liberating nucleic acid bases without loss. With sufficient formic acid in a small enough tube, as described below, however, there was no advantage in replacing the air with a non-oxidizing atmosphere.

The increased volume of formic acid was found to lead also to slightly improved yields of other bases, especially guanine, from DNA. When ox-spleen DNA was hydrolysed under these conditions total base recovery was equivalent to 98–100% of the P (Table 6). It appears that the use of an insufficient volume of formic acid has been one reason for low total recoveries of bases from DNA in some previous analyses (e.g. Wyatt, 1951*b*; Chargaff, Lipshitz, Green & Hodes, 1951; Lalland, Overend & Webb, 1952).

It will be noted that the formic acid referred to above is 88% (Merck reagent), although in previous work 98–100% (A.R.) has been used. When the two grades were tested simultaneously, no significant difference was found in yields of the bases from DNA. A similar dependence of yield upon volume was found with 98% as with 88% formic acid, and the former produced a darker-coloured hydrolysate suggesting greater degradation of the deoxypentose.

The procedure ultimately adopted for base analysis of phage DNA is as follows. Virus (1·5 mg.) or virus DNA (0·7 mg.) is weighed into a Pyrex glass tube of 6 mm. internal diameter. Formic acid (0·5 ml. of 88%) is added, and the tube is sealed off about 20 mm. above the surface of the liquid and heated to 175° for 30 min. The tube is opened, the hydrolysate evaporated to dryness under reduced pressure at a temperature not exceeding about 75°, and the residue redissolved in 25 µl. of N-HCl. Two 8 µl. portions are taken for chromatography, and two 2 µl. portions for P estimation.

Elementary analyses

P was estimated in the virus and nucleic acid preparations by the method of Bergold & Pister (1948), and N by a micro-Kjeldahl procedure. For the elementary analyses on purified 5-hydroxymethylcytosine we are indebted to the analytical laboratory of the Research Division of Sharp & Dohme, Inc.

Determinations of deoxyribonucleic acid bases in viruses and other materials

Since it was possible that some of the virus preparations might contain RNA, the early chromatograms of phages hydrolysed with $HClO_4$ (which breaks down RNA as well as DNA to free bases) were examined for possible uracil. The ratio of uracil to thymine should give an approximate indication of the ratio of RNA to DNA. Maximal values for uracil, calculated from the absorption at 260 mµ. of the appropriate areas of chromatograms, ranged from 1·9 to 2·8% of the thymine for seven preparations of T4r⁺, T4r, T6r⁺ and T6r; in no case did the absorption curve exhibit a peak at this wavelength, so that the absorption must have been largely due to other degradation products from the virus, and any actual uracil was considerably less than the figures mentioned.

The relative proportions of the DNA bases found in the r and r⁺ mutants of coliphages T2, T4 and T6 are shown in Table 3. The small difference in apparent thymine content of whole viruses compared with isolated DNA has no biological significance, but is due to interference by breakdown products of the non-DNA moiety of the virus. The very small differences in results from the various virus strains are not significant, but fall within the range of experimental error. Because of the lability of 5-hydroxymethylcytosine it is probable that even under the conditions of hydrolysis finally adopted a small amount of this base is lost. The results in Table 2 suggest that this may be of the order of 4%, and the figures presented may require corresponding correction.

From the ratios of total recovered bases to total P (Table 3) it is evident that in the best virus preparations the P may be fully accounted for as nucleotides. The possible presence of 1 or 2% of non-DNA P is not excluded, however,

Table 3. *Composition of phages T2, T4 and T6 with respect to DNA bases, phosphorus and nitrogen*

Material analysed	No. of preparations	No. of analyses	Mean estimated content of bases (moles/100 moles) Adenine	Thymine	Guanine	5-Hydroxymethylcytosine	Total bases/P (moles/g. atom)	N/P (atomic ratio)
T2r⁺ DNA	2	3	32·5	32·6	18·2	16·7	0·97	—
T2r DNA	1	3	32·4	32·4	18·3	17·0	0·98	—
T6r⁺ DNA	2	10	32·5	32·5	18·3	16·7	0·99	3·85
T2r⁺ virus	1	4	32·0	33·3	18·0	16·8	0·99	7·0
T2r virus	1	4	32·3	33·4	17·6	16·7	0·95	7·0
T4r⁺ virus	1	3	32·3	33·1	18·3	16·3	0·96	6·9
T4r virus	2	3	32·2	33·5	18·0	16·3	0·94	6·7
T6r⁺ virus	3	2	32·5	33·5	17·8	16·3	0·99	7·2
T6r virus	2	3	32·3	33·4	17·7	16·6	0·88	6·7

because base estimation in the presence of protein may err slightly on the high side. In other preparations some non-DNA P is indicated: this is generally greater in the r than in the r^+ mutants, and in two preparations of T6r averaged some 10% of the total P. This is in general agreement with the results of an earlier study, which indicated that the non-DNA P is associated with contaminants of host origin (Cohen & Arbogast, 1950*a*).

As additional characterization of the viruses, N/P ratios are included in Table 3. These are very similar for the different preparations, tending, as would be expected, to be lower in those containing extraneous P. In spite of some earlier evidence that T4 may contain less DNA per infectious unit than T2 and T6 (Cohen & Arbogast, 1950*a*), no significant difference is evident in the N/P ratios of the three viruses. Our values agree well with the mean value of 6·83 calculated from the data of Herriott & Barlow (1952) for T2.

The discovery of 5-hydroxymethylcytosine in these viruses naturally raised the question whether this substance might occur elsewhere. The biological materials listed in Table 4 were examined. Formic acid hydrolysates were subjected to chromatography in two solvents, as previously described for estimation of 5-methylcytosine (Wyatt, 1951*a*), so that maximal figures for the contents of hydroxymethylcytosine relative to cytosine were obtained. In no case did the eluate of the appropriate region of the chromatogram have the characteristic absorption spectrum of hydroxymethylcytosine, so that no evidence was obtained for the presence of this substance in any of these materials. Where only a small amount of material containing a low proportion of DNA was available, the figures for maximal hydroxymethylcytosine are necessarily raised by interfering substances.

Table 4. *Maximal contents of 5-hydroxymethylcytosine, calculated from the ultraviolet absorption of chromatogram eluates, in various biological materials*

Material examined	Wt. hydrolysed (mg.)	Maximal 5-hydroxymethylcytosine as percentage of cytosine (mol.prop.)
Dried *Esch. coli*, strain B	270	0·2
Crude DNA from *Esch. coli**	—	0·6
Ox-spleen DNA	18	0·2
Phage T5	2·1	1·0
Phage T7†	3·8	2·7
Crude DNA from phage T7‡	18	0·5
Polyhedral virus	10·6	2·4
DNA from vaccinia virus	1·2	0·6
Meningo-pneumonitis virus	14·8	3·9

* Isolated by the method of Smith & Wyatt (1951) from 400 mg. of dried bacteria.
† Preparation of Dr Kozloff.
‡ Isolated from the preparation of Dr Spizizen.

By a similar analysis of phage T6r^+, using 45 mg. of virus, its maximal content of cytosine was limited to 0·2% of its hydroxymethylcytosine.

During the course of this investigation, DNA preparations were obtained from phage T5 and from vaccinia virus of sufficient apparent purity to justify our reporting the molar ratios of the bases found in them. These are shown in Table 5, along with a new analysis of ox-spleen DNA. This was the preparation previously described and analysed (Wyatt, 1951*b*), which is apparently identical in composition with calf-thymus DNA, and the present analysis is considered to be more accurate because of better recoveries resulting from the use of a higher proportion of formic acid in hydrolysis. Quantitative analyses were also performed on phage T7; however, since there was evidence of interference by materials other than DNA, we do not report the figures and merely remark that the results indicated a closer approximation to equimolar proportions of the four bases than are in the DNA of phage T5.

Table 5. *Base composition of DNA's estimated after hydrolysis in 400–700 times their weight of 88% formic acid*

Source of DNA	No. of analyses	Moles/100 moles estimated bases				Total bases/P (moles/g. atom)
		Adenine	Thymine	Guanine	Cytosine	
Ox spleen	4	27·9	27·3	22·7	20·8*	1·00
Phage T5	2	30·3	30·8	19·5	19·5	0·93
Vaccinia virus	1	29·5	29·9	20·6	20·0	1·00

* 1·3 moles of 5-methylcytosine are also present (Wyatt, 1951*b*); this base was not estimated in the present analyses

DISCUSSION

It is necessary to reconcile certain earlier results with those now presented. The report of cytosine as a component of phage T2 by Smith & Wyatt (1951) was based on a limited amount of material, and when chromatograms were obtained identical in appearance with those of DNA from other sources it was assumed without further critical examination that spots having equal R_F values represented the same substances. The mistaken recognition of deoxycytidylic acid by Weed & Cohen (1951) resulted from the following coincidence. When the products of hydrolysis of phage T6 DNA in N hydrochloric acid were separated on paper chromatograms, a band was resolved having both the R_F value and the ultraviolet-absorption maximum (278 mμ. in 0·01 N hydrochloric acid) of deoxycytidylic acid, which it was therefore assumed to be. Recent re-examination of this fraction by Dr Weed (personal communication), however, has resulted in its resolution on ion-exchange columns into two components: one with absorption maximum at 282 mμ. at pH 2, representing deoxy-5-hydroxymethylcytidylic acid, and the other, with a maximum 274 mμ., as yet unidentified. When combined, these products had an absorption spectrum close to that of deoxycytidylic acid. The conclusions drawn from these studies, however, are not altered by the substitution of hydroxymethylcytosine for cytosine.

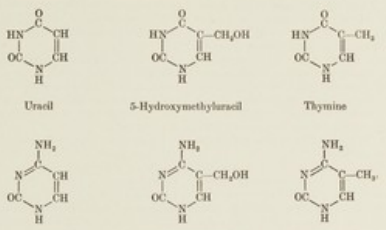
Marshak (1951) missed hydroxymethylcytosine because of his use of perchloric acid for hydrolysis along with a chromatogram solvent system in which it happens to migrate together with guanine. This accounts for the anomalous absorption spectrum for guanine which he reported.

In spite of the considerable evidence that DNA may play a specific role in the transmission of hereditary characters, we were unable to demonstrate any difference in the composition of the DNA of the r and r^+ mutants of phages T2, T4 and T6. This confirms the inference drawn from similar analyses on a number of insect viruses (Wyatt, 1952*b*) that genetic difference is not necessarily accompanied by a detectable quantitative difference in DNA composition.

A common pattern has been noted in the composition of DNA from many sources: the molar ratios (adenine)/(thymine) and (guanine)/(cytosine + 5-methylcytosine) are relatively constant and close to unity (Chargaff, 1951; Wyatt, 1952*a*). The same regularities are seen to be valid with DNA from phage T5 and from vaccinia virus, and also with DNA of phages T2, T4 and T6 except that here cytosine is replaced by 5-hydroxymethylcytosine. Whether these near-unity ratios actually signify equal numbers of the corresponding nucleotides in the molecule is as yet uncertain. The present studies, however, have served to emphasize how quantitative errors can result from small differences in experimental conditions and purity of materials, and it is our experience that successive improvements in technique have resulted in bringing the observed ratios successively closer to unity. One is tempted to speculate that regular structural association of nucleotides of adenine with those of thymine and of guanine with those of cytosine (or its derivatives) in the DNA molecule requires that they be equal in number. There is as yet, however, no direct evidence for such a theory.*

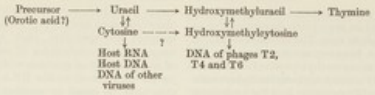
The occurrence of 5-hydroxymethylcytosine as a major constituent of the nucleic acid of a virus, none of which could be found in the host cells, presents problems of fundamental importance for the chemistry of virus production. Although discussion must at present remain largely speculative, certain possibilities may be pointed out.

We are concerned with the following pyrimidine bases:



The metabolic pathways for pyrimidines appear generally to involve their ribosides and deoxyribosides rather than the free bases, and preliminary experiments by one of us (S.S.C.) indicate that this probably is the case in *Esch. coli*. In the rat, Reichard & Estborn (1951) have demonstrated that deoxycytidine can be utilized for production of thymidine, but not vice versa. Elwyn & Sprinson (1950) have implicated the β-carbon of serine as a source of the 5-methyl group of thymine, which is evidently synthesized by methylation of a preformed pyrimidine ring. Since serine cleaves to formaldehyde, we may question whether methyl-group synthesis from serine may not involve an initial hydroxymethylation followed by reduction. If this is so, 5-hydroxymethylpyrimidines (or their deoxyribosides) could be normal metabolites, inter-

* Since this was written, a structure for DNA involving such specific pairing of nucleotides has been proposed by Watson & Crick (1953).



mediary in the synthesis of 5-methylpyrimidines, and a scheme (as above) may be tentatively proposed.

Folic acid has been implicated in the synthesis of the 5-methyl group of thymine (Goldthwait & Bendich, 1952), and it may be with the proposed hydroxymethylation step that it is concerned. This suggests a possible explanation for the inhibition of growth of phages T2, T4 and T6 by sulphanilamide under conditions (including the presence of thymine) which do permit multiplication of phages T1, T3 and T7 (Rutten, Winkler & de Haan, 1950). If the drug interferes with hydroxymethylation, so long as thymine is provided, only the growth of viruses requiring a hydroxymethylpyrimidine will be inhibited. It was with this in mind that we examined the virus of meningo-pneumonitis, which is also sensitive to sulphonamides. It does not contain hydroxymethylcytosine, hence its inhibition must be otherwise accounted for, and may be due to another function of folic acid. The theory with respect to phages T2, T4 and T6 is neither supported nor necessarily invalidated.

The proposal that 5-hydroxymethylcytosine can be derived from cytosine is supported by recent tracer experiments with *Esch. coli* infected with phage T6 (Weed & Cohen, unpublished). Two routes are possible, however: (*a*) the hydroxymethyl substituent may be added directly to cytosine (or its deoxyriboside), or (*b*) cytosine may be deaminated to uracil, which would be hydroxymethylated and then aminated to produce hydroxymethylcytosine. In the latter case 5-hydroxymethyluracil would be a precursor both of thymine and of hydroxymethylcytosine, and the last compound could be an abnormal end product produced only during growth of one of the viruses requiring it.

In either case the problem arises as to whether the virus provides the enzyme for the terminal step in the synthesis of the deoxyriboside of hydroxymethylcytosine or whether the host contains an enzyme which is not normally functioning to produce significant amounts of this compound. Preliminary data suggest that neither the pyrimidine nor its deoxyriboside is a normal intermediate in pyrimidine metabolism (Cohen, unpublished) and the mechanism which permits the new compound to be produced during some types of virus growth is a subject for continued investigation.

To what extent can this special pyrimidine contribute to an explanation of the changes in nucleic acid metabolism observed during virus multiplication? A reorganization of DNA synthesis appears to be a general phenomenon in virus-infected bacteria, and cannot depend on any unusual pyrimidine base. Total inhibition of RNA synthesis, however, has as yet been clearly demonstrated only with phages T2, T4 and T6, and may possibly be related to their content of hydroxymethylcytosine. If this substance, or the virus containing it, were to block, in the above scheme, either the production or the utilization of cytosine, a shunt of all nucleic acid synthesis into production of virus DNA, as is observed, would be the result. Just how such inhibition might be caused, however, we cannot at present say.

We have also noted (unpublished experiments) that the DNA of these viruses is more resistant to the action of deoxyribonuclease and phosphatase than is thymus DNA. This may possibly give the virus nucleic acid a selective advantage in the infected cell.

Finally, we may note that the presence of a different pyrimidine base provides a marker by which virus DNA may be distinguished from host DNA. This will be of use in determining, for example, whether virus DNA or its components are being synthesized during the first few minutes of infection of *Esch. coli* B by T2, when the net DNA content of the cell remains constant.

SUMMARY

1. The deoxyribonucleic acids of bacteriophages T2, T4 and T6 of *Esch. coli* contain no cytosine. Instead, they contain a hitherto unrecognized pyrimidine which has been isolated in crystalline form and found to be identical with synthetic 5-hydroxymethylcytosine.

2. The ultraviolet-absorption characteristics of 5-hydroxymethylcytosine are described, and those of its deamination product, 5-hydroxymethyluracil, partially described.

3. 5-Hydroxymethylcytosine is lost during hydrolysis of deoxyribonucleic acid with concentrated perchloric acid at 100°, but can be recovered almost quantitatively after hydrolysis with formic acid at 175°, provided that a sufficient volume is used. Yields of other bases, especially guanine, are also somewhat affected by the proportion of formic acid used.

4. The quantitative purine and pyrimidine composition of deoxyribonucleic acid from phages T2, T4 and T6, each in r and r⁺ mutants, has been determined, and no differences could be detected among these viruses. The molar ratios adenine/thymine and guanine/5-hydroxymethylcytosine are close to unity and the ratio (adenine + thymine)/(guanine + 5-hydroxymethylcytosine) is 1·8.

5. No 5-hydroxymethylcytosine could be detected in any of the following materials: dried cells of *Esch. coli*, ox-spleen deoxyribonucleic acid, phages T5 and T7, an insect polyhedral virus, vaccinia virus and meningo-pneumonitis virus. Quantitative analyses of deoxyribonucleic acid bases from phage T5 and from vaccinia virus are reported.

6. The possible significance of 5-hydroxymethylcytosine, which may be a component peculiar to certain viruses, is discussed in relation to the metabolism of pyrimidine bases and of nucleic acids during virus infection.

We are indebted to a number of persons, as acknowledged in the text, for gifts of materials, and are especially grateful to Dr J. M. Sprague, Dr Bettylee Hampil and other members of the staff of Sharp & Dohme Inc., Research Division, for generous co-operation with the use of their facilities. The work of one of us (S. S. C.) was conducted under a grant from the Commonwealth Fund. This paper is Contribution no. 94, Division of Forest Biology, Science Service, Department of Agriculture, Ottawa, Canada.

REFERENCES

Bergold, G. (1947). *Z. Naturf.* **2b**, 122.
Bergold, G. H. (1953). *Advanc. Virus Res.* **1**, 91.
Bergold, G. & Pister, L. (1948). *Z. Naturf.* **3b**, 406.
Chargaff, E. (1951). *Fed. Proc.* **10**, 654.
Chargaff, E., Lipshitz, R., Green, C. & Hodes, M. E. (1951). *J. biol. Chem.* **192**, 223.
Cohen, S. S. (1947). *Cold Spr. Harb. Symp. quant. Biol.* **12**, 35.
Cohen, S. S. (1951). *Bact. Rev.* **15**, 131.
Cohen, S. S. & Arbogast, R. (1950*a*). *J. exp. Med.* **91**, 607.
Cohen, S. S. & Arbogast, R. (1950*b*). *J. exp. Med.* **91**, 619.
Davidson, D. & Bandisch, O. (1925). *J. biol. Chem.* **64**, 619.
Elwyn, D. & Sprinson, D. B. (1950). *J. Amer. chem. Soc.* **72**, 2317.
Goldthwait, D. A. & Bendich, A. (1952). *J. biol. Chem.* **196**, 841.
Herriott, R. M. & Barlow, J. L. (1952). *J. gen. Physiol.* **36**, 17.
Hershey, A. D. & Chase, M. (1952). *J. gen. Physiol.* **36**, 39.
Hitchings, G. H., Elion, G. B., Falco, E. A. & Russell, P. B. (1949). *J. biol. Chem.* **177**, 357.
Hoagland, C. L., Smadel, J. E. & Rivers, T. M. (1940). *J. exp. Med.* **71**, 737.
Laland, S. G., Overend, W. G. & Webb, M. (1952). *J. chem. Soc.* p. 3224.
Litzinger, A. & Johnson, T. B. (1936). *J. Amer. chem. Soc.* **58**, 1936.
Markham, R. (1953). *The Nature of Virus Multiplication*, ed. P. Fildes & W. E. van Heyningen, Cambridge University Press.
Marshak, A. (1951). *Proc. nat. Acad. Sci., Wash.*, **37**, 299.
Marshak, A. & Vogel, H. J. (1951). *J. biol. Chem.* **189**, 597.
Putnam, F. W. (1952). *Exp. Cell Res. Suppl.* **2**, 345.
Reichard, P. & Estborn, B. (1951). *J. biol. Chem.* **188**, 839.
Rutten, F. J., Winkler, K. C. & de Haan, P. G. (1950). *Brit. J. exp. Path.* **31**, 369.
Siminovitch, L. & Rapkine, S. (1952). *Biochim. biophys. Acta*, **9**, 478.
Smith, J. D. & Wyatt, G. R. (1951). *Biochem. J.* **49**, 144.
Tamm, C., Hodes, M. E. & Chargaff, E. (1952). *J. biol. Chem.* **195**, 49.
Watson, J. D. & Crick, F. H. C. (1953). *Nature, Lond.*, **171**, 737.
Weed, L. L. & Cohen, S. S. (1951). *J. biol. Chem.* **192**, 693.
Wyatt, G. R. (1951*a*). *Biochem. J.* **48**, 481.
Wyatt, G. R. (1951*b*). *Biochem. J.* **48**, 484.
Wyatt, G. R. (1952*a*). *Exp. Cell Res. Suppl.* **2**, 201.
Wyatt, G. R. (1952*b*). *J. gen. Physiol.* **36**, 201.
Wyatt, G. R. & Cohen, S. S. (1952). *Nature, Lond.*, **170**, 1072.
Wyatt, G. R. & Cohen, S. S. (1953). *Ann. Inst. Pasteur*, **84**, 143.

EXTRAIT DES

ANNALES

DE

L'INSTITUT PASTEUR

(Janvier 1953. — Tome 84.)

THE BASES OF THE DEOXYRIBONUCLEIC ACIDS OF T2, T4, AND T6 BACTERIOPHAGES

BY

G. R. WYATT and Seymour S. COHEN

MASSON ET Cie, ÉDITEURS

Libraires de l'Académie de Médecine

120, Boulevard Saint-Germain

PARIS

Imprimé avec le périodique « *Annales de l'Institut Pasteur* »
N° d'ordre 1704. — (Extrait Janvier 1953. — Tome 84, p. 443.)

THE BASES OF THE DEOXYRIBONUCLEIC ACIDS OF T2, T4, AND T6 BACTERIOPHAGES

by G. R. WYATT, and Seymour S. COHEN (1).

(The Laboratory of Insect Pathology, Sault Ste. Marie, Ontario, and the Department of Pediatrics and the Department of Physiological Chemistry, University of Pennsylvania, Philadelphia, Pa.)

In correlating the possible specificity of deoxyribonucleic acids (DNA) with chemical structure, it is of interest to compare the nucleic acids of genetically related and unrelated viruses. This has been studied by analysing the base compositions of the genetically related *r* and *r+* strains of T2, T4, and T6 bacteriophages. The T2, T4, and T6 viruses are serologically related and may be considered to be genetically more distantly related than the one-step mutants, *r* and *r+*, within each group. Chemical and metabolic properties of these viruses have been described [**1**, **2**].

Conflicting data have appeared on the base composition of these viruses. Uracil has not been observed, confirming the absence of ribose nucleic acid. Adenine, guanine, and thymine have been found without any difficulty. Cytosine, however, although present in *E. coli*, was not found in analyses of T2 and T4 in which the ratio of recovered bases to P was low (ca. 0.7) [**3**, **4**]. However, other workers reported the presence of cytosine in T2 [**5**], and its deoxyribonucleotide in T6 [**6**]. These discrepancies now appear to be resolved by the discovery of a new base in T2, T4, and T6, which closely resembles cytosine, for which it was previously mistaken. The new base is quite labile, and for this reason, it was not found by other investigators. The new base will be referred to as C_x, since its structure, although suspected, has not yet been determined (2).

Perchloric acid hydrolysis of phage or phage DNA largely

(1) The work of one investigator (S. S. C.) was conducted under a grant from the Commonwealth Fund.

(2) *Note added in proof.* C_x has been identified as 5-hydroxymethylcytosine.

destroys C_x under conditions in which base recovery from thymus or rickettsial DNA is 96-100 p. 100. Destruction of C_x is minimized by the use of 88 p. 100 formic acid at 175° for 30 minutes. The recovery of C_x from intact phage is slightly better than that from isolated DNA. Formic acid hydrolysis of whole viruses followed by paper chromatography of the bases [7] has led to the analyses presented in table I. Among possible reasons for the

TABLE I. — **Base composition of virus DNA** (moles p. 100 moles total nucleotide).

	T2r+	T4r+	T6r+
Adenine	33.6	33.0	33.6
Guanine	18.1	18.3	18.0
Thymine	35.6	36.0	36.0
C_x	12.7	12.5	12.3
Cytosine	0	0	0
5 methyl cytosine	0	0	0
Ratio $\frac{\text{base recovery}}{\text{P}}$	0.90	0.90	0.91

relatively low base recovery, equivalent to about 90 p. 100 of the P content, are 1) the lability of C_x, and 2) the use of the molecular extinction coefficient of cytosine for C_x, which may prove to be unjustified.

Repeated analyses have failed to reveal significant differences between the *r* and *r*+ mutants nor indeed among the 6 viruses. Although these results show an apparent similarity of the 6 nucleic acids in 6 genetically distinct viruses, numerous possibilities remain of isomeric and population differences not demonstrable by the analytical methods now in use.

C_x has properties very similar to cytosine but may be separated from it by paper chromatography. In figure 1, it can be seen that in isopropanol-water mixtures, C_x has a higher R_f value than cytosine in solvents rich in water, and a lower R_f value than cytosine in those of low water content. This hydrophilic property also distinguishes C_x from 5-methyl cytosine. The relative positions taken by C_x and cytosine, however, are independent of the pH of the solvent. Examination of hydrolysates by chromatographic methods has revealed the maximal cytosine content of T6r+ to be 0.2 p. 100 of the C_x content. Conversely, the maximal C_x content of *E. coli* DNA and thymus DNA were found to be 0.6 and 0.2 p. 100 respectively of the cytosine contents.

C_x contains an amino group which may be removed by nitrous acid. Deaminated C_x then exhibits chromatographic properties

bearing a relation to those of uracil comparable to the relation of the original C_x to cytosine. The ultraviolet absorption spectra of these compounds were examined, and the maxima at various pH's, presented in table II, distinguish C_x from a nucleoside, a 5-OH derivative, or a 4-methyl derivative.

TABLE II. — **Ultraviolet absorption maxima.**

	pH 1 mμ	pH 7-8 mμ	pH 13 mμ
C_x	278	269	284, 230
Cytosine	275	267	280, 223
5-methyl cytosine	283	274	287
4-methyl cytosine	276	268	277
Cytosine	280	271	272
Desoxycytidine	280	272	272
Deaminated C_x	261	261	285
Uracil	259	259	243
Thymine	264	264	288
5-OH uracil	278	238, 239	304
4-methyl uracil	261	261	276

A new nucleotide has now been isolated from virus hydrolysates which contains a base which appears to be C_x [8]. The nucleotide migrates similarly on paper as deoxycytidylic acid but may be separated from the latter and other contaminants of the fraction on ion exchange resins.

Thus the host, *E coli* strain B, contains the pyrimidine, cytosine, which is absent from the T2, T4, or T6 viruses. Conversely, these viruses contain a new base, C_x, which is absent from the host and which seems to bear a structural similarity to cytosine. Weed and Cohen prepared pyrimidine-labeled *E. coli* by growth in the presence of labeled orotic acid [6], and infected the host in the absence of label in the medium. Virus nucleotides were isolated and the fraction containing the C_x nucleotide was heavily labeled. It appears possible that in virus synthesis host cytosine is converted to C_x. If the infected cell is no longer capable of synthesizing cytosine, it is conceivable that this alone would account for the inability of the host to continue synthesizing many host constituents. In this light, it becomes of interest to analyse a lysogenic system for the presence of C_x.

Studies are being continued on the structure and origin of C_x and its relation to other problems of virus synthesis. It is evident that if the existence of a unique virus constituent extends to animal virus systems as well, this may prove of great interest in the chemotherapy of virus infections.

BIBLIOGRAPHY

[1] S. S. Cohen and R. Arbogast. *J. exp. Med.*, 1950, **91**, 707.
[2] S. S. Cohen and R. Arbogast. *J. exp. Med.*, 1950, **91**, 619.
[3] E. Chargaff. Personal communication.
[4] A. Marshak. *Proceed. Nat. Acad. Sci.*, 1951, **37**, 229.
[5] J. D. Smith and G. R. Wyatt. *Biochem. J.*, 1951, **49**, 144.
[6] L. L. Weed and S. S. Cohen. *J. biol. Chem.*, 1951, **192**, 693.
[7] G. R. Wyatt. *Biochem. J.*, 1951, **48**, 584.
[8] L. L. Weed. Personal communication.

Imprimé par l'Ancne Imple de la Cour d'Appel, 1, rue Cassette, à Paris (France).
(50038)

A NEW PYRIMIDINE BASE FROM BACTERIOPHAGE NUCLEIC ACIDS

By Dr. G. R. WYATT and Dr. S. S. COHEN

(*Reprinted from Nature, Vol.* 170, *p.* 1072, *December* 20, 1952)

(*Reprinted from Nature, Vol.* 170, *p.* 1072, *December* 20, 1952)

A New Pyrimidine Base from Bacteriophage Nucleic Acids*

THE pyrimidine cytosine[1] and its deoxynucleotide[2] have been reported as constituents of the deoxyribonucleic acids of coliphages *T2* and *T6* respectively, but in another analysis of *T2*[3] no cytosine could be found. We have now resolved this discrepancy by isolation from *T*-even bacteriophages of a new pyrimidine base, identified as 5-hydroxymethylcytosine, which was previously mistaken for cytosine or missed. A nucleotide has now also been isolated which gives rise to this base on hydrolysis[4].

The nucleic acid bases liberated by hydrolysis of whole viruses or virus deoxyribonucleic acid were separated on paper chromatograms using isopropanol – water – hydrochloric acid[3] as the solvent. When formic acid (88 per cent at 175° C. for 30 min.) was used for hydrolysis, a substance was obtained having the same R_F value in this system as cytosine but a slightly different ultra-violet absorption spectrum. Perchloric acid, which gives good yields of the other nucleic acid bases, largely destroys this substance.

Its structure was suggested to us by its spectral and chromatographic characteristics. The ultraviolet absorption maxima were close to those of cytosine and 5-methylcytosine, having parallel shifts with change in *p*H :

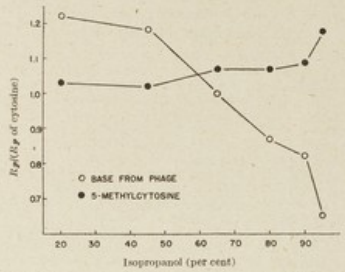
	Wave-lengths of maxima, mμ		
	0·1 *N* hydrochloric acid	*p*H 7·4	0·1 *N* sodium hydroxide
Cytosine	275	267	280
5-Methylcytosine	283	274	287
Base from phage	279	269·5	283·5

The relative positions of cytosine and the new substance on chromatograms were independent of the *p*H of the solvent, suggesting similar ionizable groups. The effect of the water content of the solvent on R_F values, however, illustrated in the accompanying graph, showed the relatively hydrophilic nature of the new base.

By treatment with nitrous acid, the base was deaminated to a product which was related in spectral and chromatographic properties to uracil in a similar way as the parent substance was to cytosine.

These properties suggested 5-hydroxymethylcytosine as a possible structure. A specimen of the base,

* Contribution No. 59, Division of Forest Biology, Science Service, Department of Agriculture, Ottawa, Canada.



isolated by a combination of chromatographic and chemical procedures from a large preparation of *T6r*$^+$ bacteriophage (for which we are indebted to Dr. J. Spizizen, of Sharp and Dohme, Inc.), had an elementary composition compatible with this structure (analyses after drying *in vacuo* at room temperature: calc. for $C_6H_7O_2N_3.\frac{1}{2}H_2O$: C, 40·00; H, 5·37; found: C, 40·23; H, 5·35 per cent. Analyses after drying over phosphorus pentoxide at 100° C.: calc. for $C_6H_7O_2N_3$: N, 29·78; found: N, 29·33 per cent). A sample of 5-hydroxymethylcytosine, kindly synthesized for us by Dr. C. S. Miller, of Sharp and Dohme, Inc., had ultra-violet spectra and chromatographic behaviour identical with those of the natural substance. On heating, both natural and synthetic products decomposed without melting. We conclude that the base found in the *T*-even bacteriophages is 5-hydroxymethylcytosine.

By hydrolysis of whole viruses in formic acid, the following analyses were obtained:

	Moles per 100 moles total nucleotide				Base recovery / *P*
	Adenine	Thymine	Guanine	Hydroxymethylcytosine	
T2r$^+$	33·2	35·2	17·9	13·6	0·91
T4r$^+$	32·7	35·7	18·1	13·5	0·91
T6r$^+$	33·2	35·6	17·8	13·3	0·92

Neither cytosine, 5-methylcytosine, nor uracil could be detected. The values for 5-hydroxymethylcytosine are based on the extinction coefficient in 0·1 *N* hydrochloric acid (E = 9,700); they may be subject to correction for loss in hydrolysis. Repeated analyses of the *r* and *r*$^+$ mutants of *T2*, *T4* and *T6* phages did not indicate any significant differences in composition of the deoxyribonucleic acid.

Careful examination of deoxyribonucleic acid from thymus and from *E. coli* and whole dried *E. coli* failed to reveal any of the new pyrimidine in these materials, and it appears that it may be a unique virus constituent. This would then provide the first demonstration of a virus containing a chemical unit qualitatively different from those of its host.

These results will be reported more fully elsewhere. This work was done while one of us (S. S. C.) was in receipt of a grant from the Commonwealth Fund.

G. R. Wyatt
Laboratory of Insect Pathology,
Sault Ste. Marie, Ontario.

S. S. Cohen
Department of Pediatrics,
Children's Hospital of Philadelphia,
and
Department of Physiological Chemistry,
University of Pennsylvania,
Philadelphia. Sept. 3.

[1] Smith, J. D., and Wyatt, G. R., *Biochem. J.*, **49**, 144 (1951).
[2] Weed, L. L., and Cohen, S. S., *J. Biol. Chem.*, **192**, 693 (1951).
[3] Marshak, A., *Proc. U.S. Nat. Acad. Sci.*, **37**, 299 (1951).
[4] Weed, L. L. (personal communication).
[5] Wyatt, G. R., *Biochem. J.*, **48**, 584 (1951).

Printed in Great Britain by Fisher, Knight & Co., Ltd., St. Albans

THE NUCLEIC ACIDS OF SOME INSECT VIRUSES

By G. R. WYATT

[Reprinted from THE JOURNAL OF GENERAL PHYSIOLOGY, November 20, 1952,
Vol. 36, No. 2, pp. 201–205]

Printed in U.S.A.

[Reprinted from THE JOURNAL OF GENERAL PHYSIOLOGY, November 20, 1952,
Vol. 36, No. 2, pp. 201–205]
Printed in U.S.A.

THE NUCLEIC ACIDS OF SOME INSECT VIRUSES*

BY G. R. WYATT

(*From the Laboratory of Insect Pathology, Sault Ste. Marie, Ontario, Canada*)

(Received for publication, July 1, 1952)

Analyses of desoxyribonucleic acids (DNA) from various sources have led to the conclusions that (1) the proportions of the purine and pyrimidine bases are characteristic of the species source, and (2) there is a regular pattern of composition whereby the ratios between certain bases remain relatively constant (5, 10). In the hope of establishing the generality of these observations and clarifying their significance it is worthwhile to continue accumulating data on the composition of DNA from various sources. The range of composition, greater among viruses than higher organisms, makes the former especially suitable for such studies, and analyses are now described of the nucleic acids of eleven insect viruses. Preliminary results on some of these have already been reported (8, 10).

EXPERIMENTAL

Preparation of the Viruses.—Viruses were isolated from purified inclusion bodies (polyhedra or capsules) by the procedures of Bergold (1, 3). Sixty to 100 mg. of inclusion bodies,[1] yielding 1.5 to 3.5 mg. of virus, were used for each analysis. For hydrolysis, each preparation was sedimented and dried in a 6 × 50 mm. glass culture tube.

Hydrolysis.—The nucleic acid bases were liberated from whole viruses with 70 per cent perchloric acid (7). Since some variability was noted in the yield of thymine, tests were carried out with purified ox spleen DNA to determine optimal conditions. Loss of thymine was found to result from temperatures above 100°C. (about 15 per cent is lost in 1 hour at 110°). Excess $HClO_4$ is also destructive: using 70 per cent $HClO_4$ at 100°C. for 1 hour, constant yields were obtained with 8 to 15 µl. $HClO_4$ per mg. DNA, about 2.5 per cent loss with 20 µl. per mg., and 10 per cent loss with 30 µl. per mg. Destruction was lessened, however, by the presence of protein, which appears to have a protective action.

The results in Table I show that under the conditions selected the presence of protein does not alter the apparent yield of any base from ox spleen DNA

* Contribution No. 38, Division of Forest Biology, Science Service, Department of Agriculture, Ottawa, Canada.

[1] Thanks are due to Dr. G. H. Bergold, Dr. F. T. Bird, and Dr. E. A. Steinhaus for provision of purified polyhedra and virus-killed insects of several species.

by more than 3 per cent, and that recoveries are not diminished by prolonging hydrolysis to 2 hours.

The procedure adopted was to add 3 to 4 µl. of 70 per cent $HClO_4$ for each milligram of dry virus and heat to 100°C. for 2 hours. After dilution of the hydrolysate with water to 25 µl., and grinding with a glass rod to suspend the insoluble residue (which tends to adsorb phosphorus), two 8 µl. portions were taken for chromatography, and two 2 µl. portions for phosphorus estimation (4).

Estimation of the Bases.—Published procedures (9) using paper chromatography and estimation in the Beckman spectrophotometer were followed. By elution of spots in 4 ml. 0.1 N HCl with mechanical shaking, and sufficient attention to exact correspondence of paper blanks to unknowns, good results could be obtained with less than 5 µg. of purine or pyrimidine per spot.

TABLE I

Yield of Bases from Ox Spleen DNA after Hydrolysis in $HClO_4$ (70 Per Cent, 15 µl. per Mg. NA) at 100°

Values are means from three experiments.

Material analyzed	Hydrolysis time	Moles per 100 moles total bases*				Percentage of total P accounted for by corresponding nucleotides	Percentage of total N accounted for
		Adenine	Thymine	Guanine	Cytosine		
	hrs.						
Purified DNA	1	28.4	26.4	23.1	20.8	95	97
" "	2	28.6	26.1	23.3	20.8	98	97
DNA + albumin (15:85)	1	28.0	27.1	22.8	20.8	99	—
" "	2	28.0	27.0	22.8	21.0	100	—

* Allowing 1.3 for methylcytosine (10), not estimated in these experiments.

RESULTS AND DISCUSSION

The proportions of the nucleic acid bases in the eleven viruses are listed in Table II. Neither uracil nor 5-methylcytosine could be detected. The amount of DNA estimated by summing the bases corresponds to 10 to 15 per cent of the various viruses, but these figures are only approximate since the procedure was not designed to refer results to dry weight. The fraction of total phosphorus accounted for by the estimated nucleotides averaged 88 per cent: the difference between this figure and 100 per cent, obtained with spleen DNA by the same procedure, may point to the presence of some non-nucleotide phosphorus in the virus or to different behavior of the two nucleic acids under hydrolytic conditions.

The viruses are listed (Table II) in order of increasing proportion of adenine. Their thymine contents then also form, within experimental error, an ascending series, while the values for both guanine and cytosine assume descending sequence. The ratios of adenine to thymine, and guanine to cytosine, are

TABLE II

Purine and Pyrimidine Composition of DNA of Insect Viruses

	Host species	Host order and family	No. of analyses*	Moles per 100 moles total bases				Percentage of total P accounted for
				Adenine	Thymine	Guanine	Cytosine	
Polyhedral viruses	*Porthetria dispar* (L.), gypsy moth	Lepidoptera Lymantriidae	4	21.2 ± 0.15‡	20.05 ± 0.18	30.5 ± 0.11	28.25 ± 0.09	92
	Lymantria monacha L., nun moth	"	1	24.6	23.8	26.8	24.7	—
	Choristoneura fumiferana (Clem.), spruce budworm	Tortricidae	3	24.8 ± 0.12	24.0 ± 0.09	26.7 ± 0.14	24.5 ± 0.14	86
	Psychoda sarisata Schrk.	Geometridae	2	26.7	25.7	24.4	23.2	87
	Malacosoma americanum (F.), Eastern tent caterpillar	Lasiocampidae	3	29.2 ± 0.22	28.0 ± 0.34	22.5 ± 0.19	20.2 ± 0.11	93
	Malacosoma disstria Hbn., forest tent caterpillar	"	3	29.2 ± 0.23	28.5 ± 0.37	21.9 ± 0.19	20.3 ± 0.07	86
	Bombyx mori L., silkworm	Bombycidae	3	29.3 ± 0.25	28.0 ± 0.33	22.5 ± 0.05	20.2 ± 0.13	88
	Colias philodice eurytheme Bdvl., alfalfa butterfly	Pieridae	4	29.9 ± 0.35	27.6 ± 0.08	22.4 ± 0.15	20.1 ± 0.22	90
	Neodiprion sertifer (Geoffr.), pine sawfly	Hymenoptera Tenthredinidae	2	32.3	30.3	19.5	17.9	—
Capsule viruses	*Cacoecia murinana* Hb.	Lepidoptera Tortricidae	3	32.1 ± 0.14	30.5 ± 0.44	19.7 ± 0.35	17.8 ± 0.29	84
	Choristoneura fumiferana (Clem.), spruce budworm	"	1	32.8	32.4	18.4	16.4	—

* Independent analyses performed on different preparations of virus.

‡ Mean value and its standard error.

nearly constant for all, whereas the ratio of adenine plus thymine to guanine plus cytosine (AT:GC ratio) indicates the position of a virus in the series (Table III).

The scale of AT:GC ratios appears to be a discontinuous, stepwise distribution, with groups of viruses having similar values. Within each of these groups no significant differences (by the *t* test) have been found. This outcome was scarcely expected in view of the results from animal DNA, which suggested a unique composition for each species, and those from tobacco mosaic virus

TABLE III

Molar Proportions in Insect Virus Nucleic Acids

		Molar ratios in DNA			
	Virus host	Adenine Thymine	Guanine Cytosine	Purines Pyrimidines	AT/GC
Polyhedral viruses	*P. dispar*	1.06	1.08	1.07	0.71
	L. monacha	1.03	1.08	1.06	0.94
	C. fumiferana	1.03	1.09	1.06	0.95
	P. seriata	1.04	1.05	1.04	1.10
	M. americanum	1.04	1.11	1.07	1.34
	M. disstria	1.02	1.08	1.05	1.36
	B. mori	1.04	1.11	1.07	1.34
	C. P. eurytheme	1.08	1.11	1.09	1.35
	N. serifer	1.07	1.09	1.07	1.67
Capsule viruses	*C. murinana*	1.05	1.11	1.07	1.67
	C. fumiferana	1.01	1.12	1.05	1.87
	Ox spleen DNA*	1.04	1.02	1.03	1.22

* Included for comparison; ratios are taken from the analysis of DNA in the presence of protein, and methylcytosine has been added to cytosine.

(6), which indicate that even related virus strains can differ in nucleic acid composition.

The viruses which have been analyzed are all (with the possible exception of those from *Malacosoma americanum* and *M. disstria*) distinct on the basis of morphology, inclusion body type, and host specificity (2). Some relationship between these properties and the AT:GC ratios may be inferred since the two capsule viruses fall adjacent to one another in the series, and so do the polyhedral viruses whose hosts belong to one family. No direct parallel can exist, however, between biological relationship and DNA composition, since the groups having equal AT:GC ratios include viruses as unlike as any examined. Clearly, any genetic specificity carried by DNA must reside in more subtle differences than can be revealed by simple measurement of the base ratios.

The independence of host and virus nucleic acids, already demonstrated for certain bacteriophages (8), is confirmed by the unlike DNA compositions of a polyhedral virus and a capsule virus from one insect species, *C. fumiferana*. Several attempts to prepare DNA from the host caterpillars led to only very small and impure yields; approximate analyses, however, of DNA from *B. mori* and *L. monacha* gave ratios similar to those of other animals and unlike those of the viruses.

SUMMARY

Purine and pyrimidine bases have been estimated from the desoxyribonucleic acids of eleven insect viruses. Their proportions vary in the different species in a balanced way so that the molar ratios adenine:thymine and guanine:cytosine are constant and close to unity, whereas adenine + thymine:guanine + cytosine ranges from 0.71 to 1.87. This ratio is identical for some biologically dissimilar viruses, and no general parallelism is evident between DNA composition and biological relationship. Two different viruses from one host have distinct DNA's.

REFERENCES

1. Bergold, G., *Z. Naturforsch.*, 1947, **2b**, 122.
2. Bergold, G. H., *Ann. New York Acad. Sc.*, 1952, in press.
3. Bergold, G. H., *Adv. Virus Research*, 1953, in press.
4. Bergold, G., and Pister, L., *Z. Naturforsch.*, 1948, **3b**, 406.
5. Chargaff, E., *Fed. Proc.*, 1951, **10**, 664.
6. Markham, R., and Smith, J. D., *Biochem. J.*, 1950, **46**, 513.
7. Marshak, A., and Vogel, H. J., *J. Biol. Chem.*, 1951, **189**, 597.
8. Smith, J. D., and Wyatt, G. R., *Biochem. J.*, 1951, **49**, 144.
9. Wyatt, G. R., *Biochem. J.*, 1951, **48**, 584.
10. Wyatt, G. R., *Exp. Cell Research*, 1952, suppl. 2, 201.

THE JOURNAL OF GENERAL PHYSIOLOGY
VOLUME 36, NUMBER 2, NOVEMBER 20, 1952

CONTENTS

WAVERLY PRESS, INC.
BALTIMORE

(*Reprinted from Nature, Vol.* 170, *p.* 846, *November* 15, 1952)

Nucleic Acids of Rickettsiæ*

Smith and Stoker[1] have noted that the deoxyribonucleic acid of *Q*-fever rickettsiæ (*R. burneti*) grown in embryonated eggs is closely similar in composition to that of chick embryos, the only significant difference being lack of 5-methyl cytosine from the former. To account for this similarity, they mention the possibility that rickettsiæ may incorporate nucleic acid directly from their host. This hypothesis could be tested, they point out, by analysis of the same organism grown in a different host.

We have recently analysed the deoxyribonucleic acid of another rickettsia from the same host. Deoxyribonucleic acid was isolated[2] from epidemic typhus rickettsiæ (*R. prowazeki*) grown in chick yolk-sac endothelium, and the purine and pyrimidine bases estimated[3] after hydrolysis in perchloric acid (70 per cent, 1 hr. at 100° C.). The estimated bases were equivalent to 96 per cent of the deoxyribonucleic acid phosphorus. No methyl cytosine could be detected. The ratios of the bases found (mean of two analyses), along with those reported for *R. burneti*, are as follows :

	Moles per cent			
	Adenine	Thymine	Guanine	Cytosine
R. burneti (ref. 1)	29·5	26·0	22·5	22·0
R. prowazeki	35·7	31·8	17·1	15·4

The deoxyribonucleic acid of *R. prowazeki* clearly differs in composition from those of *R. burneti* and of the host chick embryos. In the *E. coli*-phage system, the deoxyribonucleic acids of host and virus differ markedly[4], and it has been concluded from tracer studies[5] that host nucleic acid is utilized by the parasite only in the form of breakdown products. It is likely that any host nucleic acid used in synthesis of rickettsiæ would also be first extensively degraded.

A common pattern in the composition of deoxyribonucleic acid, first pointed out by Chargaff[6], has apparently wide validity. The molar ratios (adenine) : (thymine), and (guanine) : (cytosine (plus methyl cytosine when present)) have nearly constant values close to unity, whereas the ratio (adenine + thymine) : (guanine + cytosine) is characteristic of the source of the nucleic acid. The rickettsiæ are no exception :

	A : T	G : C	(A + T) : (G + C)
R. burneti	1·13	1·02	1·25
R. prowazeki	1·12	1·11	2·08

* Contribution No. 51, Division of Forest Biology, Science Service, Department of Agriculture, Ottawa, Canada.

If, for as yet obscure reasons, nearly constant ratios of adenine to thymine and guanine to cytosine are the rule in deoxyribonucleic acids, this would greatly increase the probability of two distinct deoxyribonucleic acids having like composition by chance. This may account for the similarity of the *R. burneti* and chick nucleic acids.

G. R. Wyatt

Laboratory of Insect Pathology,
Sault Ste. Marie,
Ontario.

S. S. Cohen

Department of Pediatrics,
Children's Hospital of Philadelphia,
and the
Department of Physiological Chemistry,
University of Pennsylvania,
Philadelphia.

April 25.

[1] Smith, J. D., and Stoker, M. G. P., *Brit. J. Exp. Path.*, **32**, 433 (1951).
[2] Cohen, S. S., *J. Immunol.*, **65**, 475 (1950).
[3] Wyatt, G. R., *Biochem. J.*, **48**, 584 (1951).
[4] Smith, J. D., and Wyatt, G. R., *Biochem. J.*, **49**, 144 (1951).
[5] Weed, L. L., and Cohen, S. S., *J. Biol. Chem.*, **192**, 693 (1951). Kozloff, L. M., *J. Biol. Chem.*, **194**, 95 (1952).
[6] Chargaff, E., *Fed. Proc.*, **10**, 654 (1951).

Printed in Great Britain by Fisher, Knight & Co., Ltd., St. Albans.

(Expt. Cell Res. Suppl. 2, 201 (1952))

SPECIFICITY IN THE COMPOSITION OF NUCLEIC ACIDS*

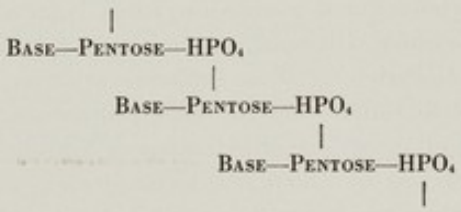
G. R. WYATT

Laboratory of Insect Pathology, Sault Ste. Marie, Ontario

NUCLEIC ACID STRUCTURE RECONSIDERED

Among the major chemical cell components, the nucleic acids remain today probably the least understood. This may seem surprising, since the basic units in the structure of nucleic acids are fewer in number than those of proteins, carbohydrates, or lipids, but this apparent simplicity was deceptive and led to a falsely simplified view of these substances which tended to inhibit enquiry for many years. Recently it has come to be realized that nucleic acids are very complex molecules possessing considerable biological specificity, the physico-chemical bases of which are currently being sought.

Below are listed the known constituent units of nucleic acids, with a diagrammatic indication of their linkage into polynucleotide chains:



	Ribonucleic acids (RNA)	Desoxyribonucleic acids (DNA)
Pentose	Ribose	Desoxyribose
Purines	Adenine	Adenine
	Guanine	Guanine
Pyrimidines	Cytosine	Cytosine
	Uracil	Thymine
		5-Methylcytosine

Most of these constituent substances have been recognized for many years. It seems likely that D-ribose and D-2-desoxyribose are the exclusive carbohydrate components of RNA and DNA respectively, for all nucleic acids give color reactions

*Contribution No. 26, Division of Forest Biology, Science Service, Department of Agriculture, Ottawa, Canada.

that partially characterize these sugars, and no evidence has yet been brought forward for the presence of any others. They have been unequivocally identified, however, in nucleic acids from only a few sources (14). The recent recognition of 5-methylcytosine as a new pyrimidine component of many desoxypentose nucleic acids (45) should serve as a warning against any complacent view that no new components remain to be discovered.

There are a number of possible bases for specificity in nucleic acids. The nature of the internucleotide linkages and the structure and configuration of polynucleotide chains are not yet definitely settled for any nucleic acid, so their possible contribution to specificity, great though it may be, cannot be usefully discussed at this stage. The order of sequence of different nucleotides in the chains is undoubtedly important and is being investigated, by analysis of products of partial hydrolysis, in several laboratories (28, 47). Our present knowledge, however, is only sufficient to emphasize the complexity of the problem. A simpler, if more superficial, approach to the differentiation of nucleic acids is by estimation of the relative amounts of the several purine and pyrimidine bases they contain. Results from this type of analysis, particularly with DNA, form the subject of this paper.

As soon as the major purine and pyrimidine bases from nucleic acids had been identified, early in this century, attempts were made to determine their quantitative relations. Chemical fractionation of these substances is not easy, as anyone will agree who has attempted to isolate them by chemical means, or who has analyzed commercial preparations of them. Nevertheless, some reasonably complete analyses were obtained, two of the best of which for "thymonucleic acid" (which may have come from thymus or fish sperm) are shown in Table 1. The proportions of the

TABLE 1
Composition of Thymonucleic Acid by Chemical Isolation of the Bases

Base	% Dry wt	Molar ratios*	% Dry wt	Molar ratios*
	Steudel (39)		Levene and Mandel (24)	
Adenine	10.68	1.30	8.17	0.98
Guanine	9.01	0.98	9.15	0.98
Cytosine	4.26	0.63	7.00	1.02
Thymine	8.33	1.09	8.00	1.02

*Molar ratios are calculated to total 4.00.

bases were not far from equimolecular. Similar results, with uracil instead of thymine, were obtained from nucleic acids of the pentose type from plants (wheat germ (33) and yeast (22)). As the researchers recognized the inaccuracy of the methods and were influenced by concepts borrowed from the chemistry of small molecules, they assumed that their results indicated equimolecular proportions of

four bases. This idea grew into the "tetranucleotide hypothesis" which postulated groups of two purine and two pyrimidine nucleotides, in fixed order, as the basic units of nucleic acid structure. It was implicit that there could be only one RNA and one DNA. This simplified view is dogmatically stated in Levene and Bass' monograph (23), and dominated thought on nucleic acids for many years.

More recently, physical measurements have shown that nucleic acids can have very high molecular weights—of the order of 10^6 in the case of thymus DNA. At the same time, biological evidence has accumulated suggesting that they play roles in heredity and protein synthesis which seem to require some specificity of structure. These developments engendered a new critical attitude toward their chemistry, and the tetranucleotide theory was questioned, in the first instance by Gulland (16). Evidence obtained by various methods during the past few years has all run counter to the old theory.

The road to more accurate knowledge of the composition of nucleic acids was finally opened by the development of two modern techniques, partition chromatography and ultraviolet spectrophotometry. These now provide simple and precise micro-methods for the analysis of both types of nucleic acid. The feasibility of separating nucleic acid derivatives by chromatography on filter paper was first demonstrated by Vischer and Chargaff (40) and by Hotchkiss (18), and during the past three years there has followed the publication of quantitative methods and results from a number of laboratories. All investigators agree that the composition of nucleic acids is not consistent with the tetranucleotide hypothesis, and most have recognized quantitative differences between nucleic acids from different sources.

QUANTITATIVE ANALYSIS OF NUCLEIC ACIDS

This symposium is not the place for detailed descriptions of technique, but a few remarks on the merits of various procedures may be of general interest. It has been found that truly quantitative separations of nucleic acid derivatives can be obtained by chromatography on filter paper, and for routine analytical work this method has the advantages of simplicity and sensitivity over the use of columns. Ion-exchange columns have proved better for special problems and preparative purposes. For detection of the spots on paper chromatograms, one can utilize either their fluorescence (17) or their absorption in the ultraviolet (26); the latter technique has been found very convenient, since it is sensitive to less than 1 μg of purine base and provides a permanent photographic record of each chromatogram.

As a solvent for the separation of the free bases on paper chromatograms, aqueous isopropanol (65% in 2.0 *N* HCl) (46) is very satisfactory, since the two purines and four pyrimidines commonly found in nucleic acids can all be resolved in a single one-dimensional run. The strong acid is helpful by increasing the solubility of guanine, of which only small amounts will migrate quantitatively in neutral or

weakly acidic or basic solvents. This mixture also resolves the nucleosides and some, but not all, of the nucleotides.

The nucleic acid derivatives eluted from chromatograms are estimated by their ultraviolet absorption in the quartz spectrophotometer. The method is sensitive and precise, but depends on accurate knowledge of extinction coefficients, and there is some disagreement in the literature on these (18, 34, 36, 41, 46). Agreement on these coefficients, re-determined with substances of established purity, would permit better comparison of results from different laboratories.

The problem of obtaining quantitative hydrolysis of nucleic acids has been troublesome, and many methods have been used. The purine bases are easily liberated from both RNA and DNA with dilute mineral acid, but splitting of the pyrimidine nucleosides requires much more vigorous treatment which frequently leads to deaminated or otherwise degraded products (36, 42). Separate hydrolyses for the purines and the pyrimidines have often been used, but since several methods are now known which permit estimation of all the bases from a single hydrolysate, only these will be discussed.

From RNA, free purines along with pyrimidine nucleotides can be obtained by hydrolysis for 1 hr in *N* HCl at 100° C (36). This method is not applicable to DNA, but the latter can be split completely to free bases by pure formic acid at 175° C for 30 min (46). A treatment yielding free bases from both types of nucleic acid has been described by Marshak and Vogel (31), using 70% $HClO_4$ at 100° for 1 hr. This method gives consistent results, although the yields of the bases differ significantly from those obtained by hydrolysis with formic acid (Table 2), and this must be allowed for when results from the two methods are compared.

TABLE 2
Composition of Bovine DNA According to Various Authors

Author	Molar ratios (total = 4.00)					% of P accounted for
	A	T	G	C	MC	
Chargaff et al. (10)	1.20	1.09	0.96	0.74	—	90
Marshak & Vogel (31)	1.19	1.15	0.83	0.82	—	96
Daly, Allfrey, & Mirsky (13)	1.11	1.15	0.93	0.81	—	99
Hurst, Marko, & Butler (20)	1.08	1.14	0.905	0.835	0.046	96
Wyatt (HCOOH)	1.13	1.12	0.86	0.85	0.053	91
Wyatt ($HClO_4$)	1.14	1.04	0.92	0.84	0.05	99

While enzymic methods of hydrolysis have usually given rise to complex mixtures of products, improved knowledge of purified enzymes is increasing their usefulness, and a quantitative enzymic method is being used in Butler's laboratory (19, 20). A specific phosphodiesterase obtained from venom of Russell's viper enables re-

duction of nucleic acid to mononucleotides, which can be separated on anion exchange columns and estimated by means of their phosphorus content. Because of its freedom from strong chemical treatments and its dependence on phosphorus estimations rather than ultraviolet absorption measurements, this method provides a useful check on the other techniques.

By suitable micro-techniques in conjunction with paper chromatography, purines, pyrimidines, and phosphorus can be estimated from a sample containing less than 0.5 mg of nucleic acid. In the work to be described on insect viruses, the standard errors of the means of triplicate analyses, each using 2 to 3 mg of virus, containing 0.3 to 0.5 mg of DNA, have regularly been less than $\pm$ 1.5%. This is, of course, a measure of the reproducibility of the method, not of its absolute accuracy.

The proportions of the purines and pyrimidines in DNA from bovine tissues as estimated in several laboratories are listed in Table 2. In all cases, the ratios of the several bases have been calculated to an arbitrary total of 4.00; this procedure has no logical basis now that it is known that not all nucleic acids contain only four bases, but it will be used throughout this paper for sake of uniformity. Where DNA from more than one bovine tissue has been analyzed in one laboratory, mean results are given, since no significant differences have been noted among tissues of one species. The figure for percentage of total phosphorus accounted for, on the assumption of a 1:1 molar relationship between bases and phosphorus in the nucleic acid, gives some measure of the completeness of an analysis. It may be pointed out, however, that technical errors can as easily make this figure erroneously high as erroneously low, and that equimolarity of phosphorus and bases in all nucleic acids is not definitely established. Considering the diversity of the methods used, agreement between the results of different authors is fairly close.

Analyses of RNA's in several laboratories have shown that these too depart widely from tetranucleotide structure and contain different proportions of the purines and pyrimidines according to their source (9, 36, 42). Constitutional differences have even been found between RNA's of different strains of tobacco mosaic virus (27) and between those of nuclei and cytoplasm of one tissue (29).

PURINE AND PYRIMIDINE COMPOSITION OF DESOXYPENTOSE NUCLEIC ACIDS

I should now like to present some of my own results on the composition of nucleic acids of the desoxypentose type.

One of the first observations during these studies was the recognition of the pyrimidine 5-methylcytosine as a new component of certain DNA's (44, 46). This pyrimidine had been reported in 1925 by Johnson and Coghill (21) as a constituent of tuberculinic acid; hence, when in 1949 an extra substance was observed in DNA from herring sperm, 5-methylcytosine seemed a likely identity for it. On isolation,

the new substance proved indeed to have the properties of 5-methylcytosine and to be identical with the synthetic compound. It was present in nucleic acid from many sources, but not a trace could be found in tubercle bacilli. Its absence from tubercle bacillus DNA has also been determined by Vischer, Zamenhof and Chargaff (43) so Johnson and Coghill's original claim, which provided the first clue to its identity, was apparently erroneous.

Methylcytosine has been found in amounts much smaller than those of the other bases in the DNA of all higher animals and plants yet examined, but not in bacteria and viruses. A complete picture of its distribution, however, depends upon analysis of many more species from different phyla of organisms. The constancy of the amounts found in preparations from a given source, and the fact that these are unaltered by treatment of the nucleic acid with NaOH or ribonuclease, indicate that it is a true component of DNA. The corresponding nucleotide (12) and nucleoside (15) have now also been recognized.

The molar ratios of the bases in DNA from eight animal sources and wheat germ, estimated after hydrolysis with formic acid, are shown in Table 3. It is

TABLE 3
Composition of DNA's from Animal Sources and Wheat Germ. Formic Acid Hydrolysis. (Wyatt (46))

Source	Molar ratios (total = 4.00)				
	A	T	G	C	MC
Calf thymus	1.13	1.11	0.85	0.85	0.052
Beef spleen	1.13	1.12	0.85	0.85	0.054
Bull sperm	1.15	1.09	0.89	0.83	0.052
Ram sperm	1.15	1.09	0.88	0.84	0.039
Rat bone marrow	1.15	1.14	0.86	0.82	0.044
Herring sperm	1.11	1.10	0.89	0.83	0.075
Locusta migratoria	1.17	1.17	0.82	0.83	0.008
Echinus esculentus sperm	1.24	1.18	0.78	0.74	0.071
Wheat germ	1.05	1.08	0.94	0.69	0.23

obvious that there is great similarity among the different preparations. The differences in proportions of the major bases between one nucleic acid and another are in some cases significant but in others, no greater than the experimental error. The figure for methylcytosine, however, is significantly different for every species, and is the same only in the three nucleic acids from bovine tissues, thymus, spleen, and sperm. This corroborates the conclusion of Chargaff (7), based on analyses in which methylcytosine was not included, that DNA composition is a species character, differing for different species of organisms, but not for different tissues of one species. It is also worth noting that there is greater similarity of DNA composition among the various vertebrates than there is between these and two invertebrates

examined, a locust and a sea urchin, and that the DNA of wheat differs strikingly from the others in its high content of methylcytosine. There is thus a suggestion that the degree of similarity in the DNA of various species may bear some relation to their phylogenetic proximity.

A large group of viruses pathogenic to insects provides an especially convenient subject for nucleic acid analysis. These viruses cause the diseases of caterpillars and sawfly larvae known as polyhedral diseases and capsule or granule diseases (3, 4, 5) because of the abundant crystalline inclusion bodies produced during infection. These inclusion bodies can be purified in relatively large quantities from the diseased insects. By dissolving them in very dilute alkali the viruses can be liberated and isolated in a state of purity surpassed probably only by the crystallizable viruses of plants.

A great many species of insects are subject to these diseases, over 100 having already been recorded (38), and in general each seems to have its own specific virus, so that the viruses are here designated by the names of their hosts. Those that have been analyzed contain about 15% of nucleic acid, all of the desoxypentose type (6, 37) and extremely little lipid or non-nucleic acid carbohydrate. They can satisfactorily be hydrolyzed whole with perchloric acid for purine-pyrimidine analysis, as the protein degradation products do not interfere with estimation of the bases.

The compositions of the nucleic acids of several of these viruses are shown in Table 4. Methylcytosine was not found. It is notable that there are very wide differences in nucleic acid composition between some species, while others are almost if not quite identical, and the identities are not found where they might be expected on biological grounds. Thus, there is no significant difference between

TABLE 4
Composition of DNA's of Insect Viruses.
Perchloric Acid Hydrolysis

Virus designation*	Number of analyses	Molar ratios (total = 4.00)			
		A	T	G	C
Pd	3	0.845 ± 0.009**	0.80 ± 0.009	1.225 ± 0.003	1.13 ± 0.005
Bm	3	1.17 ± 0.010	1.12 ± 0.013	0.90 ± 0.002	0.81 ± 0.005
Cpe	3	1.185 ± 0.013	1.105 ± 0.003	0.90 ± 0.008	0.81 ± 0.008
Ma	1	1.18	1.11	0.91	0.80
Cm	2	1.285	1.235	0.775	0.70
Ns	2	1.29	1.21	0.785	0.715

* Pd = Gypsy moth (*Porthetria dispar* L.) Polyhedral virus.
Bm = Silkworm (*Bombyx mori L.*) Polyhedral virus.
Cpe = Alfalfa butterfly (*Colias philodice eurytheme* Boisd.) Polyhedral virus.
Ma = Tent caterpillar (*Malacosoma americanum* Fabr.) Polyhedral virus.
Cm = European budworm (*Cacoecia murinana* Hb.) Capsule virus.
Ns = Pine sawfly (*Neodiprion sertifer* Geoff.). Polyhedral virus.
**Mean and its standard error.

the polyhedral viruses of the silkworm (a moth) and the alfalfa butterfly, insects which fall in different sub-orders, and a provisional analysis of tent caterpillar polyhedral virus suggests that it too has the same composition. On the other hand, the virus of gypsy moth differs very greatly from those of silkworm and tent caterpillar, insects falling in different families of the same sub-order. Again, there is great similarity in nucleic acid between the capsule virus of the European budworm and the polyhedral virus of the pine sawfly; here not only are different orders of insects involved, but the viruses are characterized by the production of different types of inclusion body.

These comparisons tend to cast doubt on any idea of a direct parallelism between DNA composition and taxonomic relationships of organisms. The relationships of these viruses to one another, however, are not well understood, and it does not necessarily follow that they parallel those of their hosts, although among such highly adapted host-specific parasites this seems likely. Studies on the serological relationships of these viruses may help in understanding the significance of their nucleic acid composition.

Among all these animal, plant, and viral nucleic acids, however, certain regularities in the pattern of variation are evident. As first pointed out by Chargaff, Zamenhof, Brawerman, and Kerin (11), DNA's may be grouped by composition into two main types, an "AT type" having adenine and thymine the major bases, and a "GC type" in which guanine and cytosine preponderate. Animal DNA's were all of the AT type, while both types occurred among micro-organisms. Chargaff (8) has further observed a remarkable constancy in the ratios of adenine to thymine, guanine to cytosine, and total purines to total pyrimidines among DNA's from a number of sources. These generalizations are supported by the analyses described above.

The ratios between the amounts of different bases in animal and wheat germ DNA's are presented in Table 5. Cytosine and methylcytosine are grouped together because the most regular ratios are thus obtained, a tendency being noted for a greater methylcytosine content in a nucleic acid to be balanced by a reduced amount of cytosine. The ratios A:T, G:C+MC, and purines:pyrimidines are virtually constant, being in all cases nearly equal to unity. The ratio A+T:G+C+MC, on the other hand, gives a figure which, along with the methylcytosine content, distinguishes one nucleic acid from another.

The corresponding ratios for the insect virus nucleic acids have been computed in Table 6. The same regularities are here evident, even though both AT and GC types are included. It should be noted that these viruses were hydrolyzed with perchloric acid, whereas formic was used for the animal DNA's. When tested on the same preparation of nucleic acids perchloric acid gave a greater yield of guanine and a lower yield of thymine, each by about 6%, than formic acid (Table

TABLE 5
Molar Proportions in DNA's of Animals and Wheat.
Formic Acid Hydrolysis

Source	A / T	G / C + MC	Purines / pyrimidines	A + T / G + C + MC
Calf thymus	1.02	0.95	0.99	1.27
Beef spleen	1.01	0.94	0.98	1.28
Bull sperm	1.05	1.01	1.03	1.26
Ram sperm	1.05	1.00	1.03	1.27
Rat bone marrow	1.01	1.00	1.00	1.33
Herring sperm	1.01	0.98	1.00	1.23
Locust	1.00	0.98	0.99	1.41
Echinus sperm	1.05	0.96	1.01	1.52
Wheat germ	0.98	1.02	1.00	1.15

TABLE 6
Molar Proportions in DNA's of Insect Viruses.
Perchloric Acid Hydrolysis

Virus designation*	A / T	G / C	Purines / pyrimidines	A + T / G + C
Pd	1.06	1.08	1.07	0.70
Bm	1.04	1.11	1.07	1.34
Cpe	1.07	1.11	1.09	1.34
Ma	1.07	1.13	1.10	1.34
Cm	1.04	1.11	1.08	1.71
Ns	1.07	1.11	1.03	1.67

* See Table 4.

2). If allowance is made for this difference in method, the constant ratios found among the insect viruses become almost identical with those found for the animal nucleic acids, so that these relationships appear to be characteristic of a number of nucleic acids from very diverse sources. Whether the closeness of these ratios to unity actually implies equal numbers of nucleotides cannot be stated until the systematic errors of the methods are better known.

The compositions of DNA from two species of bacteria are shown in Table 7.

TABLE 7
Composition of DNA's of Two Bacteria (Smith & Wyatt (37))

Species	Hydrolysis	Molar ratios (total = 4.00)			
		A	T	G	C
M. tuberculosis	HCOOH	0.72	0.78	1.16	1.35
E. coli	$HClO_4$	0.92	1.10	0.96	1.03

In neither of these was methylcytosine found. The nucleic acids of tubercle bacilli and of *E. coli* are clearly different from each other and from any of the other nucleic acids analyzed. They still exhibit some regularities of composition similar to those observed above (Table 8), although the A:T, G:C, and purine:pyrimidine ratios are here lower than those of animal and insect virus DNA's. How much regularity there may be in these ratios among bacteria cannot be guessed from so small a sample.

TABLE 8
Molar Proportions in DNA's of Two Bacteria

Species	Hydrolysis	$\frac{A}{T}$	$\frac{G}{C}$	$\frac{\text{Purines}}{\text{pyrimidines}}$	$\frac{A+T}{G+C}$
M. tuberculosis	HCOOH	0.92	0.86	0.88	0.60
E. coli	$HClO_4$	0.84	0.93	0.88	1.01

The nucleic acids of bacteriophages are of particular interest in view of current biochemical studies on the multiplication of these viruses. Marshak (30) has recently reported the absence of cytosine from phage T_2. Smith and Wyatt (37), however, reported analyses of phages T_2 and T_5 which indicated that both contained appreciable amounts of cytosine, although it was much less than the other bases. In order to reconcile these results, a re-investigation of the nucleic acids of some bacteriophages is under way. It is possible that contamination with host DNA has caused trouble. It is at any rate clear that the DNA's of certain of these viruses are very differently composed from that of *E. coli*, a fact which precludes any direct transfer of nucleic acid from host to virus. The phage DNA's also do not seem to fit into the pattern of constant ratios described above.

A print of a chromatogram of the purines and pyrimidines of some typical DNA's and yeast RNA is shown in Plate I.

MOLECULAR NATURE OF DNA

Analytical results of the sort that has been presented naturally suggest attempts at calculation of the minimum molecular weight that can be fitted by the empirical ratios, and the numbers of nucleotides of each type in such a unit. The corresponding problem for amino-acid residues in proteins has been tackled with some success, but the difficulties are greater for nucleic acids than for proteins. This is because of the small number of different building units, and because the difficulty of interpreting physical measurements on such asymmetric particles as those of DNA makes the deduced molecular weights unreliable.

The majority of estimates from physical data indicate molecular weights of the order of 10^6 for DNA's. This corresponds to some 3000 nucleotides, or, if four kinds are present in roughly equal amounts, 700-800 of each. It is obvious that

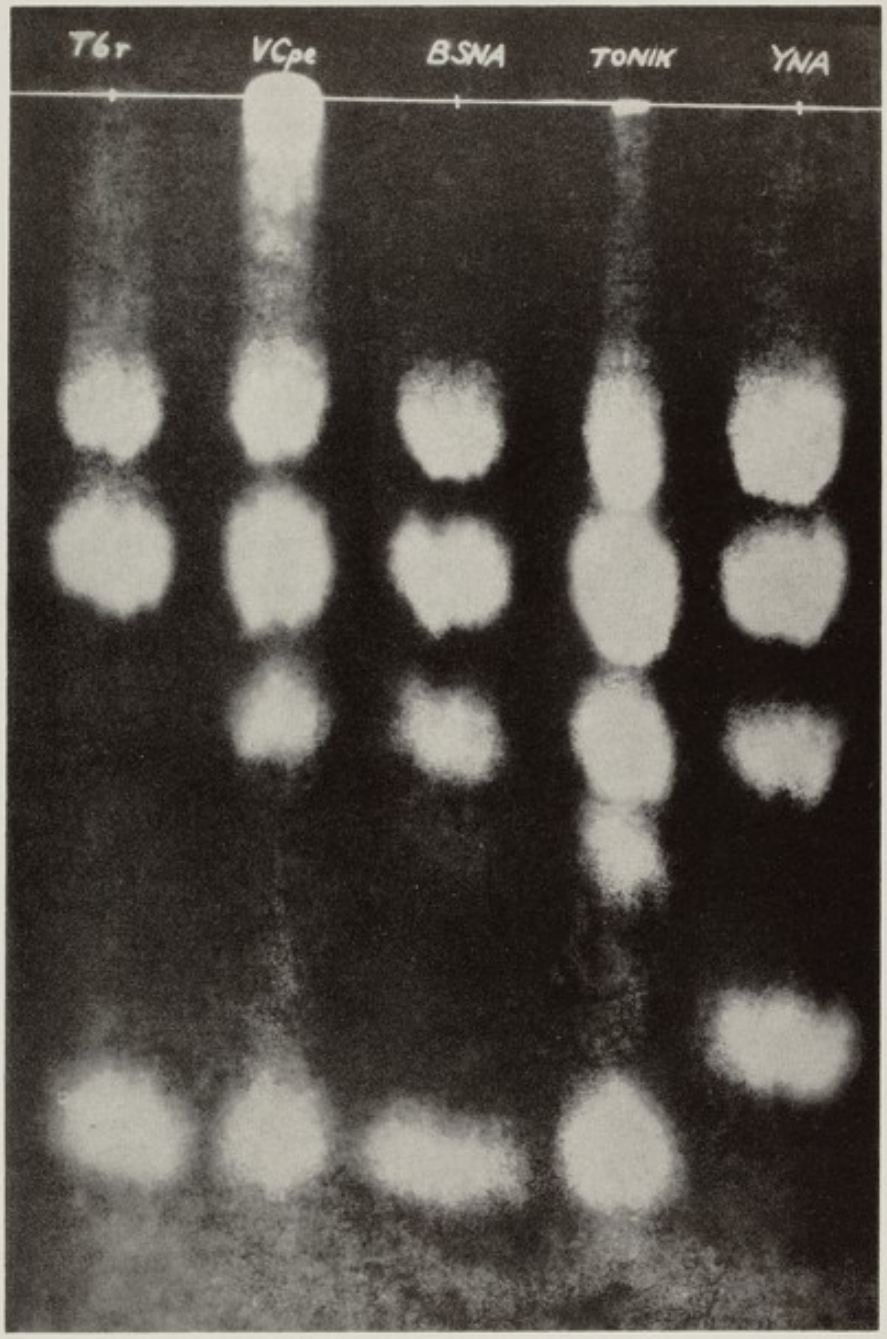


Plate I. A contact print, taken with ultraviolet light, of a paper chromatogram of the purines and pyrimidines of some typical nucleic acids. T_6r: *E. coli* bacteriophage T_6r; VCpe: alfalfa butterfly polyhedral virus; BSNA: beef spleen DNA; TONIK: wheat germ DNA; YNA: yeast RNA.
1: Guanine; 2: Adenine; 3: Cytosine; 4: 5-Methylcytosine; 5: Uracil; 6: Thymine.

a few of one kind could be replaced by another and escape detection by present methods of analysis. In such a nucleic acid, it is clearly futile to attempt to compute from the analysis a minimum molecular weight or the exact number of nucleotides of each kind in the molecule. The small content of methylcytosine in some nucleic acids, however, can provide a basis for calculation, for if it is assumed that the molecules are all of one type, there must be at least one methylcytosine nucleotide in each. Some relative nucleotide frequencies and corresponding minimum molecular weights on this basis are given in Table 9. The nucleotide numbers can be only

TABLE 9
Approximate Nucleotide Frequencies and Minimum Molecular Weights in Some DNA's, Based on the Content of Methylcytosine

Source	Relative numbers of nucleotides						Min. mol. wt.
	A	T	G	C	MC	Total	
Beef spleen	21	21	16	16	1	75	23,000
Herring sperm	15	15	12	11	1	54	17,000
Wheat germ*	9	9	8	6	2	34	11,500
Locust	146	146	102	104	1	499	155,000

* At least 2 methylcytosine nucleotides must be assumed, to obtain ratios which agree with the analysis within ± 5%.

approximate since allowance must be made for experimental error. The minimum molecular weights implied by the methylcytosine contents, although far below the molecular weights estimated for DNA, are sufficient to imply a remarkable specificity of structure. It is, indeed, difficult to imagine the synthesis of a polynucleotide chain having one residue of methylcytosine regularly in every 500, as implied by the composition of locust DNA.

An alternative hypothesis is that the molecules are not all of one type, but differ in their content of methylcytosine and presumably of the other bases also. There is no reliable evidence for the homogeneity of any DNA preparation; apparent centrifugal and electrophoretic monodispersity cannot be trusted because of boundary-sharpening effects. Singer (35) has found that a mixture of DNA's from *Escherichia coli* and bacteriophage T_2 appears homogeneous in the ultracentrifuge and during electrophoresis. In the case of RNA, there is good evidence that preparations are heterogeneous (1) and that the nucleic acid exists in different forms in the cell (2, 29). In this symposium, Dr. Bendich* has now put forward the first direct evidence of metabolic and structural heterogeneity of DNA from one tissue. It seems likely, therefore, that DNA preparations contain mixtures of molecular species, and the analyses which have been described represent merely the aggregate composition of these.

*See "Studies on the Metabolism of Nucleic Acids," p. 181.

COMPOSITION AND FUNCTION OF DNA

How can these observations on the composition of desoxypentose nucleic acids be related to present concepts regarding the function of these substances? It is generally considered that nucleic acids are essential for the multiplication of self-reproducing elements in the cell, RNA being involved in the production of cytoplasmic proteins, and DNA in the hereditary material of the nucleus. The lines of evidence which tend to link DNA with the genes have been reviewed in this symposium by Dr. Stern,* and may be summarized as:

(1) The localization of DNA in the chromosomes

(2) The metabolic stability of DNA as a cell component

(3) The production of mutations by ultraviolet irradiation, by X-rays, and by radiomimetic substances which also depolymerize DNA

(4) The production of "transformations" or mutations in certain bacteria by DNA preparations

(5) Constancy in the amount of DNA per set of chromosomes, as a species character.

To these may now be added, from the results from Chargaff's laboratory and those described in this paper:

(6) Constancy of DNA composition, as a species character.

By compiling data on the DNA content per nucleus of a large number of animal species, Mirsky and Ris (32) have demonstrated some correlation between these quantities and the evolutionary relationships of the species. The analytical results now suggest that there is also a general increase in complexity of DNA with evolutionary development of organisms, for it has been noted that the DNA's of certain bacterial viruses may be composed principally of three kinds of nucleotide, while those of other viruses and bacteria contain four, and those of some higher animals and plants have five. Many more species need to be examined, particularly from among the lower groups of animals and plants, to establish the validity of this generalization.

To carry speculation a step further, certain of these lines of evidence, in particular (3) and (4), are consistent with the idea that DNA may play a part in determining the special properties of individual genes. If this is so, each nucleus must contain a number of unique forms of DNA, for it is inconceivable that a single molecular type could possess the characteristics of all the genes of an organism. Each DNA preparation would then contain a number of molecular species perhaps even equal to the number of genes possessed by the source organism, and the analytical results would represent merely the mean composition of the total.

*See "Problems in Nuclear Chemistry and Biology," p. 1.

There is one feature of the results of purine-pyrimidine analyses that is consistent with the idea that we are dealing with the mean composition of numerous gene-specific nucleic acids. This is the observation of much greater constancy of composition among higher organisms than lower ones. It is notable, for example, that there is more similarity in DNA composition between such distantly related metazoa as locust, sea urchin, and ox than there is within a group of closely related lower organisms such as the polyhedral viruses of insects. This could follow from the greater tendency to average out among the large number of genes of higher organisms than among the relatively few genes of viruses. If, for example, the T-even bacteriophages possess 15 to 30 genetic units, as Luria and Dulbecco (25) have postulated, a single mutation might conceivably be reflected in a detectable change in the total nucleic acid composition, whereas among the thousands of genes of higher organisms a great many mutations would be required to alter the mean composition measurably.

It is difficult, however, to reconcile the complexity implied by this concept of nucleic acids with the apparent regularity connoted by the curiously constant ratios between certain of the bases. If the nucleic acids of individual genes do differ in composition, their variation must follow a fixed pattern by which these ratios are kept unchanged. We can scarcely even speculate upon how this could occur.

Also difficult to interpret is the apparently identical DNA composition of certain biologically distinct insect viruses. However, equality of composition of nucleic acids does not necessarily imply biological identity. Other possible determinants of specificity about which we know next to nothing are the order of sequence of nucleotides and the native configuration of the polynucleotide chains. The concept of gene-specific DNA's, then, can at present be put forward only as a tentative hypothesis, to stand or fall with evidence from other directions.

These analyses of the purine and pyrimidine composition of nucleic acids represent one approach, for which methods are now available, to a vast field which may prove to be the meeting point of chemistry and genetics.

References

1. Bacher, J. E. and Allen, F. W., *J. Biol. Chem.* **184**, 511 (1950)
2. Barnum, C. P. and Huseby, R. A., *Arch. Biochem.* **29**, 7 (1950)
3. Bergold, G., *Z. Naturforsch.* **2b**, 122 (1947)
4. Bergold, G., *Z. Naturforsch.* **3b**, 338 (1948)
5. Bergold, G. H., "Insect Viruses" *in* Viruses 1950. Pasadena, California, 1950
6. Bergold, G., and Pister, L., *Z. Naturforsch.* **3b**, 406 (1948)
7. Chargaff, E., *Experientia*, **6**, 201 (1950)
8. Chargaff, E., *Federation Proc.* **10**, 654 (1951)
9. Chargaff, E., Magasanik, B., Vischer, E., Green, C., Doniger, R. and Elson, D., *J. Biol. Chem.* **186**, 51 (1950)
10. Chargaff, E., Vischer, E., Doniger, R., Green, C. and Misani, F., *J. Biol Chem.* **177**, 405 (1949)

11. Chargaff, E., Zamenhof, S., Brawerman, G. and Kerin, L., *J. Am. Chem. Soc.* **72**, 3825 (1950)
12. Cohn, W. E., *J. Am. Chem. Soc.* **72**, 2811 (1950)
13. Daly, M. M., Allfrey, V. G. and Mirsky, A. E., *J. Gen. Physiol.* **33**, 497 (1950)
14. Davidson, J. N., The Biochemistry of the Nucleic Acids. Methuen, London, 1950
15. Dekker, C. A., and Elmore, D. T., *J. Chem. Soc.* 2864 (1951)
16. Gulland, J. M., Barker, G. R. and Jordan, D. O., *Ann. Rev. Biochem.* **14**, 175 (1945)
17. Holiday, E. R. and Johnson, E. A., *Nature* **163**, 216 (1949)
18. Hotchkiss, R. D., *J. Biol. Chem.* **175**, 315 (1948)
19. Hurst, R. O., Little, J. A. and Butler, G. C., *J. Biol. Chem.* **188**, 705 (1951)
20. Hurst, R. O., Marko, A. M. and Butler, G. C., personal communication, 1951
21. Johnson, T. B. and Coghill, R. D., *J. Am. Chem. Soc.* **47**, 2838 (1925)
22. Levene, P. A., *Biochem. Z.* **17**, 120 (1909)
23. Levene, P. A. and Bass, L. W., Nucleic Acids. Chem. Catalog Co., New York, 1931
24. Levene, P. A. and Mandel, J. A., *Biochem. Z.* **10**, 215 (1908)
25. Luria, S. E. and Dulbecco, R., *Genetics* **34**, 93 (1949)
26. Markham, R. and Smith, J. D., *Biochem. J.* **45**, 294 (1949)
27. Markham, R. and Smith, J. D., *Biochem. J.* **46**, 513 (1950)
28. Markham, R. and Smith, J. D., *Research* **4**, 344 (1951)
29. Marshak, A., *J. Biol. Chem.* **189**, 597 (1951)
30. Marshak, A., *Proc. Natl. Acad. Sci. U. S.* **37**, 299 (1951)
31. Marshak, A. and Vogel, H. J., *J. Biol. Chem.* **189**, 597 (1951)
32. Mirsky, A. E. and Ris, H., *J. Gen. Physiol.*, **34**, 451 (1951)
33. Osborne, T. B. and Harris, I. F., *Z. physiol. Chem.* **36**, 85 (1902)
34. Ploeser, J. McT. and Loring, H. S., *J. Biol. Chem.* **178**, 431 (1949)
35. Singer, J. S., personal communication, 1951
36. Smith, J. D. and Markham, R., *Biochem. J.* **46**, 509 (1950)
37. Smith, J. D. and Wyatt, G. R., *Biochem. J.* **49**, 144 (1951)
38. Steinhaus, E. A., Principles of Insect Pathology. McGraw-Hill, New York, 1949
39. Steudel, H., *Z. physiol. Chem.* **49**, 406 (1906)
40. Vischer, E. and Chargaff, E., *J. Biol. Chem.* **168**, 781 (1947)
41. Vischer, E. and Chargaff, E., *J. Biol. Chem.* **176**, 703 (1948)
42. Vischer, E. and Chargaff, E., *J. Biol. Chem.* **176**, 715 (1948)
43. Vischer, E., Zamenhof, S. and Chargaff, E., *J. Biol. Chem.* **177**, 429 (1949)
44. Wyatt, G. R., *Nature* **166**, 237 (1950)
45. Wyatt, G. R., *Biochem. J.* **48**, 581 (1951)
46. Wyatt, G. R., *Biochem. J.* **48**, 584 (1951)
47. Zamenhof, S. and Chargaff, E., *J. Biol. Chem.* **187**, 1 (1950)

Discussion

Oster: The idea that there are different compositions of nucleotides in nucleic acid may not be consistent with the X-ray results obtained by Dr. D. P. Riley and myself at the Royal Institution, London. We examined the X-ray "powder" diagram of highly polymerized desoxyribose nucleic acid. Nucleic acid dried over P_2O_5 shows only three broad diffraction rings in the neighborhood of 3-5 Å. A slight increase in moisture content makes no change but at 44% (by volume) of nucleic acid, there suddenly appears a magnificent display of some twenty sharp lines corresponding to spacings between about 3 and 16 Å. The spacing of 3.4 Å observed by

Astbury and Bell (*Nature* **141**, 747 (1938)) is, according to our data, actually a whole series of lines which they failed to resolve. These spacings do not change appreciably with increased water content indicating that they correspond to spacings within the molecules themselves. At higher moisture content there appear two larger spacings which vary with water content corresponding to inter-particle distances (RILEY, D. P. and OSTER, G., *Biochim. Biophys. Acta* **7**, 526 (1951)).

The sharpness of the diffraction pattern shows that the nucleic acid (with sufficient amount of water) is a substance of extreme regularity. I feel that if the regularity occurred only in the backbone of the highly polymerized chain molecule and the side groups (the purines, pyrimidines and sugar) were placed randomly along the chain, such a large number of spacings as observed would not be present. In any case, the excellent analyses carried out by Dr. Wyatt will be extremely valuable when attempts are made to try to analyze the meaning of the spacings.

Wyatt: I should like to ask Dr. Oster what types of groupings the X-ray method is capable of detecting. Can it differentiate between the nucleotides containing different bases? If not, you would get the same diffraction pattern even if you had different arrangements of nucleotides along the chain.

Oster: No. At this early stage of the analysis of the X-ray diagrams we cannot distinguish between nucleotides containing different bases. If, however, one could incorporate say, iodine, which scatters X-rays about one hundred times more than does carbon or nitrogen, into one of the bases without destroying the general arrangement of the nucleotides then one should be able, by means of X-ray diffraction, to locate the spacings between such bases. If no diffraction was forthcoming from such a procedure then I would agree that the bases were not arranged in a regular pattern along the chain.

Wyatt: I don't know enough about the nature of the chemical bonds involved to comment very much upon this but I think it is quite possible, particularly if you have a spiral structure, that the purine and pyrimidine residues are sticking out free so that they don't interfere with each other. Then you could have a regular spacing down the backbone of the chain, in spite of the differences in sequence.

Stern: The molecular models which I mentioned this morning confirm what Dr. Wyatt just said. If you construct such a model you will find that the bond angles and the distances of different groupings along the backbone are more or less independent of the sequence of the individual nucleotides. So, I agree with the inference that even if the sequence were changed, purely physical methods, like X-ray diffraction, would hardly be able to pick that up. I would like to mention that one of the principles of variation which I mentioned this morning is actually open to experimental proof: that the different DNA's in the same chromosome might differ simply by chain length. This is a problem which the polymer chemists have studied in great detail. We have very sensitive criteria for chain length measurements and for polydispersity measurements by flow birefringence for instance. A very simple experiment which I would suggest would be to isolate, by exactly the same methods, DNA from two living forms, one a very low form, where one would expect very few genes to be present, and one a highly developed form. The distribution curves for both DNA preparations could then be determined with regard to molecular weights and chain lengths. If this idea is correct there should be a positive correlation between these physical parameters and the number of different genes present in the chromosomes of the two species.

Haurowitz: If the nucleic acids consist of a chain of alternate molecules of carbohydrate and phosphoric acid, in which each carbohydrate molecule is linked to a purine or pyrimidine base, then the distribution of electric charges will be the same for all nucleic acids, and it is hard to imagine any specificity of such molecules. If, however, loops can be formed by hydrogen bonds between the phosphate groups and basic groups of the purine or pyrimidine bases, then an infinite number of specific nucleic acids is imaginable. The molecular models mentioned by Dr. K. G. Stern may be helpful in investigating this type of specificity.

Stern: There is one point which we mentioned in our earlier paper (*Yale J. Biol. and Med.* **19**, 937 (1947)) along these lines. That is, if you look at the formula of the typical tetranucleotide model you will find that the two purine bases and the two pyrimidine bases, both in DNA and in PNA, show a complementary structure at one part of the ring which makes for hydrogen bond formation. One has the CO group while the other has an NH_2 group. Now, if they vary in the sequence of purine-pyrimidine then these two complementary groups within the same polynucleotide chain are so far away from each other, a distance of over 7 Å according to Astbury's measurements, that hydrogen bonding is not very probable. But, if you rearrange them so that you have the sequence purine-pyrimidine and that the nucleotide side chains are projecting from the same side of the backbone, then these complementary groups approach each other over a distance of 3.5 Å where hydrogen bonding is just still possible. I might say that D. O. Jordan (A.C.S. meeting, San Francisco, 1949) has discussed the possibility of hydrogen bonding particularly with regard to adjacent polynucleotide chains. He feels that the fact that he gets different electrometric titration curves when he titrates forward and then titrates backwards, is due to the breakage of hydrogen bonds during titration. I think that there is a good probability of both types of bonding (i.e., of intra- and inter-molecular hydrogen bonds) in nucleic acids.

Bendich: There is recent evidence obtained by Dr. Cavalieri in our laboratory indicating that there are covalent linkages present in the nucleic acid structure, which occur between the phosphate and amino and hydroxyl groups of the bases. In addition, there is evidence for salt-type linkages, i.e., between phosphate and amino groups (Cavalieri, L. F. and Angelos, A., *J. Am. Chem. Soc.*, **72**, 4685 (1950); Cavalieri, L. F., Kerr, S. E. and Angelos, A., ibid, **73**, 2567 (1951)).

[From THE BIOCHEMICAL JOURNAL, Vol. 49, No. 2, pp. 144–148, 1951]

PRINTED IN GREAT BRITAIN

The Composition of some Microbial Deoxypentose Nucleic Acids

By J. D. SMITH and G. R. WYATT*
Agricultural Research Council Plant Virus Research Unit, Molteno Institute, University of Cambridge

(*Received* 8 *November* 1950)

Recent advances in nucleic acid chemistry, principally through the application of chromatographic techniques, have led to the abandonment of the tetranucleotide hypothesis and the discovery that both deoxypentose nucleic acids (DNA) and ribonucleic acids (RNA) show a range of compositions characteristic of their source. Because of the close association of DNA with the chromatin material of the nucleus, comparative studies on the structure of DNA from different organisms have a special biological significance. Analyses by Chargaff and co-workers have led to the conclusion that, among animals, the composition of DNA (i.e. the molar proportions of the purine and pyrimidine bases) is characteristic of the species source and constant for different tissues of one species, and that DNA from bacteria also shows characteristic differences in composition (Chargaff, 1950; Chargaff, Zamenhof, Brawerman & Kerin, 1950). Results obtained in this laboratory with improved chromatographic techniques and methods of hydrolysis, while in general agreement with these conclusions, have also led to the recognition of the pyrimidine, 5-methylcytosine, as a constituent of DNA from animals and higher plants; this is apparently lacking from nucleic acids of microbial origin (Wyatt, 1950, 1951*a*, *b*). It is apparent that in molecules so large as those of the nucleic acids many possibilities exist for structural differences which do not involve differences in nucleotide composition; nevertheless it is only through such quantitative analysis that an approach may at present be made to the comparison of different nucleic acids. We have therefore extended these analyses to include some bacterial and virus deoxypentose nucleic acids, and shall now describe the determination of the purine and pyrimidine composition of DNA from *Mycobacterium tuberculosis*, *Escherichia coli*, the two *E. coli* bacteriophages T_2 and T_5, and the virus causing polyhedral disease in gipsy-moth (*Lymantria dispar* L.) caterpillars.

It is now recognized that all bacteria which have been examined contain both RNA and DNA, and there is evidence that the DNA is confined to the bacterial nucleus (Robinow, 1945). *E. coli*, which contains about 4 % (dry wt.) of DNA, was chosen in order to permit comparison of the nucleic acids in a virus-host system. The tubercle bacillus was of particular interest because of the reported presence of 5-methylcytosine in its nucleic acid (Johnson & Coghill, 1925), although Vischer, Zamenhof & Chargaff (1949) had been unable to confirm this claim.

The bacteriophages and the polyhedral viruses of insects are convenient for nucleic acid analysis

* Present address: Laboratory of Insect Pathology, Box 490, Sault Ste Marie, Ontario, Canada.

because of their relatively high content of DNA, and are, incidentally, the only viruses which have been shown conclusively to contain only nucleic acid of the deoxypentose type. The corresponding claim for certain animal viruses (e.g. vaccinia; Hoagland, Lavin, Smadel & Rivers, 1940) has been based on the properties of nucleic acid isolated with the use of hot 5% (w/v) aqueous sodium hydroxide, which would eliminate by hydrolysis any RNA that had been present. There has been controversy over the possible presence of RNA in the bacteriophages T_2 and T_4; recently, however, Cohen & Arbogast (1950) have shown by serological methods that RNA in T_2, T_4 and T_6 preparations is associated with material antigenically related to the host and that RNA is almost certainly not a constituent part of these viruses. The virus which may be isolated from gipsy-moth polyhedral bodies has been reported to contain about 14% of DNA, sufficient to account for its total phosphorus content (Bergold, 1947, 1950; Bergold & Pister, 1948).

MATERIALS

Mycobacterium tuberculosis

Samples of dried bacterial cells of human and bovine strains were obtained through the courtesy of the Wellcome Research Laboratories.

Escherichia coli

The mutant of the strain sensitive to the T series of bacteriophages designated as B/r (Witkin, 1946) was grown on nutrient broth (Lemco, 3 g.; peptone, 10 g.; NaCl, 5 g.; distilled water, 1 l.; pH 7·5) and on a synthetic medium of the following composition: sodium lactate, 10 g.; NH_4Cl, 1 g.; K_2HPO_4, 0·7 g.; KH_2PO_4, 0·3 g.; Na_2SO_4, 0·1 g.; $MgSO_4.7H_2O$, 0·01 g.; distilled water, 1 l. Stock cultures were subcultured daily on each medium. For the nucleic acid preparations 36 or 48 hr. cultures, grown on agar slopes in Roux flasks at 37° and in the stationary phase of growth, were rapidly washed three times in 0·9% (w/v) aqueous NaCl, twice in 90% (v/v) aqueous ethanol and in ethanol, suspended in ether for 30 min., centrifuged, and dried at 105° for 6 hr.

Bacteriophages T_2 *and* T_4

The bacteriophages T_2 and T_4 (Demerec & Fano, 1945) were grown on the B/r strain of *E. coli*. Stock suspensions were titrated and kept at 4°. The usual plaque-counting method of assay was used throughout the experiments. The procedure for growth and purification of the bacteriophages followed closely that used by Putnam, Kozloff & Neil (1949) for the isolation of the bacteriophage T_4. Broth lysates were prepared by inoculating a young aerated culture at 37° with a small number of bacteriophage particles and incubating for 8–16 hr. Here several cycles of adsorption and lysis were involved. Synthetic-medium lysates were prepared by 'single-stage growth'. To an aerated culture, containing about $1–3 \times 10^8$ bacteria/ml., bacteriophage was added to give 1–5 particles/bacterium and incubation continued for about 12 hr. The bacteriophage was isolated by differential centrifugation. The lysate was centrifuged at 3000 rev./min. to remove bacterial debris, and the supernatant passed through the Sharples supercentrifuge at 24,000 rev./min. sufficiently slowly to sediment over 95% of the virus. The pellet was suspended in a small volume of 0·85% (w/v) aqueous NaCl, centrifuged at low speed (15 min. at 3000 rev./min.) to remove insoluble material, and the virus again sedimented on the Sorvall SS1 angle centrifuge at 12,000 rev./min. for 1 hr. After resuspension in saline, more debris and aggregated material were removed by low-speed centrifuging.

As the purification of bacteriophage T_6 has not been described in the literature we record here some of the characteristics of our preparations. The ultraviolet absorption spectrum in neutral solution shows a peak at 260 mμ. and trough at 240 mμ., and is very similar to the absorption curves of T_2 and T_4 bacteriophages. The preparations, both from broth and the synthetic medium, were analysed for total N and P and assayed for infective particles. These results are shown in Table 1. The quantity of N/infective particle is higher than that reported for T_2 and T_4 bacteriophages ($1·3 \times 10^{-16}$ g. N/infective particle) and corresponds with the greater diameter of T_6 particles as shown by measurements on electron micrographs (Putnam, 1950). The mean diameter of T_6 given by Putnam is 90 mμ., while that for a dry particle calculated from the ratio N/infective particle, assuming an N content of 16% and a specific volume identical with that of T_2 and T_4 (0·66), is 80–85 mμ. Considering the errors involved in the various measurements the agreement is reasonably good.

Gipsy-moth polyhedral virus

Viruses causing the polyhedral diseases of insects were first shown by Bergold (1947) to have the form of rod-shaped particles which can be liberated from the polyhedra by dissolving the surrounding 'polyhedral protein' in dilute alkali. These polyhedra are obtained from infected caterpillars, in which they are freed in great abundance on lysis of the tissues at death, and can be purified and washed by centrifugation. The diseased gipsy-moth caterpillars (*Lymantria dispar* L.) which provided polyhedra for the present work were collected from an infestation in Slovakia in June, 1949. The polyhedra were purified centrifugally, and repeatedly washed in 0·9% (w/v) aqueous NaCl until a white suspension was obtained which showed very few bodies other than polyhedra on examination with the dark-field microscope. After the final sedimentation from saline, the polyhedra were suspended in distilled water and stored at −10° until needed.

Table 1. *Characteristics of the* T_6 *bacteriophage preparations*

	g. N/infective particle ($\times 10^{-16}$)	N/P ratio (by wt.)	Extinction coefficient at 260 mμ. of a solution containing 1 g. atom N/l.
Broth preparations (mean of two preparations)	3·3	3·55	1280
Synthetic-medium preparation	3·9	3·35	1180

Virus was prepared from the purified polyhedra by the following procedure, which is essentially that of Bergold (1947). Polyhedra were suspended at a concentration of about 5 mg./ml. in a solution containing Na_2CO_3, 0·008 M, and NaCl, 0·05 M, and the mixture was allowed to stand at room temperature for about 3 hr. Examination with the dark-field microscope then showed that almost all of the polyhedra had dissolved, leaving free virus particles and empty skins or ghosts which represent the insoluble surface layer of the polyhedral bodies. The suspension was centrifuged at 6000 rev./min. for 6 min. on the Sorvall angle centrifuge to remove undissolved polyhedra, most of the skins, and any bacteria or other gross impurities. The yield of virus could be slightly increased by suspending and resedimenting this material. The supernatant liquid was then centrifuged at 9600 rev./min. for 60 min., and the virus sedimented as a translucent, bluish-white, birefringent pellet, on top of which was a less compact brownish layer consisting mainly of polyhedral skins. This impurity could be washed off, with the loss of some virus, by a carefully directed stream of water from a pipette. The virus was suspended in distilled water (about one-eighth of the volume of the original solution) and again sedimented for 60 min. at 9600 rev./min. It was then suspended in a small volume of water, and centrifuged briefly (about 2 min. at 5000 rev./min.) to remove larger particles and aggregated material.

It has not been possible to test the infectivity of such preparations in this laboratory. Examination of several of them with the electron microscope (by Dr V. E. Cosslett in the Cavendish Laboratory, Cambridge) showed that they consisted mainly of rod-shaped particles about 20×300 mμ. in size, occurring both singly and in bundles, and similar to those which Bergold has identified as the virus. Analysis of several preparations gave mean values N, 15·1% and P, 1·77% of the dry wt. (drying at 110°). The P content is rather higher than that reported by Bergold (1·33%) and appears to reflect differences in the DNA content of his preparations and ours.

METHODS

The isolation of the nucleic acids

The procedures described by Mirsky & Pollister (1946) for the isolation of deoxyribonucleoproteins from animal and plant tissues are often ineffectual with bacteria, possibly because of the impermeability of the cell wall. It has been shown in this laboratory (Wyatt, 1951*b*) that treatment with N-NaOH at 37° does not change the purine and pyrimidine composition of DNA, although RNA is completely hydrolysed to nucleotides (Schmidt & Thannhauser, 1945). This method was therefore used to prepare the nucleic acids from both bacteria and viruses. The material was incubated for 16–20 hr. at 37° in a small volume of N-NaOH. Insoluble material was removed by centrifuging, the supernatant liquid brought to pH 4 with acetic acid, and one volume of ethanol added. The precipitate, containing DNA, was collected by centrifugation, dissolved in dilute NaOH or alkaline N-NaCl, and protein removed by gel formation on shaking with a mixture of 8 parts of $CHCl_3$ and 1 part of commercial octanol (v/v) (Sevag, Lackman & Smolens, 1938). A further precipitation with acetic acid and ethanol completed the purification. Nucleic acid thus prepared was free of RNA and contained very little protein, but when derived from bacteria usually contained considerable extraneous material apparently polysaccharide in nature.

Analysis of the nucleic acids

The purine and pyrimidine composition of the nucleic acids was determined by techniques which have already been fully described (Markham & Smith, 1949; Wyatt, 1951*b*). Following hydrolysis either with pure formic acid (175° for 30 min.) or with 72% (w/w) aqueous $HClO_4$ (100° for 60 min.), the bases were separated on paper chromatograms using as the solvent 65% (v/v) aqueous *iso*propanol 2·0 N with respect to HCl. They were then eluted and estimated spectrophotometrically.

For the bacterial nucleic acids, which were contaminated with carbohydrate, hydrolysis with $HClO_4$ was found advantageous, since the use of formic acid leaves extraneous absorbing materials which interfere with chromatography. These are largely destroyed by $HClO_4$, and a cleaner chromatogram results.

With viruses such as bacteriophages and insect viruses (which contain a high proportion of nucleic acid all of one type, and little or no carbohydrate other than that of the nucleic acid) it was found that good purine-pyrimidine chromatograms can be obtained without isolation of the nucleic acid, by hydrolysis of the whole virus. Although such chromatograms are heavily loaded with amino-acids, these do not interfere appreciably with estimation of the nucleic acid bases. Tryptophan is destroyed by acid hydrolysis and thus only tyrosine has sufficiently intense ultraviolet absorption to show on the prints of the chromatograms; with the solvent used it moves in a position between cytosine and uracil, where it does not interfere. This method was used in some of the analyses of the polyhedral virus, and gave proportions of the bases similar to those found from purified virus nucleic acid.

In these analyses the information considered of interest was the relative proportions of the purine and pyrimidine components of the nucleic acids. As the reliability of the analytical procedures had already been established, N and P analyses were made only on those preparations of which sufficient was available.

Elementary analyses

N was estimated by a micro-Kjeldahl technique (Markham, 1942), and P by the colorimetric method of Allen (1940).

RESULTS

All the bacterial and virus DNA's analysed contained the purines adenine and guanine and the pyrimidines cytosine and thymine. 5-Methylcytosine could not be detected in any of these nucleic acids, even when the appropriate bands from large chromatograms were eluted, concentrated and run in a different solvent, a method which is capable of detecting a pyrimidine occurring in amounts as small as 0·05% of a nucleic acid (Wyatt, 1951*a*).

The molar proportions of the four bases in each nucleic acid are presented in Table 2. Each analysis is based on three or more replicate sets of spots on the chromatogram. The number of preparations of each nucleic acid analysed and the method of hydrolysis used are stated and the mean results given.

Table 2. *Proportions of purines and pyrimidines in deoxypentose nucleic acids from microbial sources*

Source of nucleic acid	No. of preparations analysed	Method of hydrolysis	Molar proportions calculated to total 4·00				Purines/ pyrimidines (mol. proportions)
			Adenine	Guanine	Cytosine	Thymine	
Bacteria:							
Tubercle bacilli							
Human strain	1	H.COOH	0·72	1·14	1·34	0·80	0·87
Bovine strain	1	H.COOH	0·71	1·17	1·35	0·76	0·89
Escherichia coli (mutant B/r):							
Synthetic medium	1	$HClO_4$	0·90	0·98	1·03	1·09	0·89
Broth	2	$HClO_4$	0·93	0·94	1·02	1·10	0·88
Viruses:							
E. coli phages							
T_3 synthetic medium	1	$HClO_4$	1·36	0·85	0·43	1·37	1·23
T_3 broth	1	H.COOH	1·34	0·82	0·41	1·44	1·17
T_3 broth	2	{H.COOH} * {$HClO_4$ }	1·33	0·73	0·505	1·44	1·08
Gipsy-moth polyhedral virus	4	{$HClO_4$ } * {H.COOH}	0·86 ±0·01	1·20 ±0·01	1·12 ±0·04	0·81 ±0·04	1·07

* Mean results given.

The DNA preparations from *Escherichia coli* grown on broth and on the synthetic medium do not differ significantly in composition. The same is true of DNA isolated from bacteriophage T_3 from bacteria grown on the two media. Similarly, the differences between the human and bovine strains of tubercle bacilli may probably be accounted for by systematic error, although from the few analyses that have been done, the possibility of small real differences cannot be excluded. The differences between T_3 and T_4, on the other hand, we consider too great to be accounted for by analytical error, and represent a real difference between the compositions of the two nucleic acids.

The T_3 bacteriophage preparation isolated from bacteria grown on broth was also examined for the presence of RNA. A suspension of the virus containing 0·9 mg. nitrogen was precipitated with acetic acid at pH 4, dried, and incubated with N-sodium hydroxide at 37° for 12 hr. in order to convert any RNA to mononucleotides. The solution was made acid with hydrochloric acid and the precipitate containing DNA was removed on the centrifuge. The supernatant liquid was concentrated, placed on a chromatogram, and run with the isopropanol-hydrochloric acid solvent. In this solvent the ribonucleotides from RNA give two spots, one containing uridylic acid and the other a mixture of adenylic, guanylic and cytidylic acids. Both these run clear of the purines adenine and guanine. No well defined spots were distinguishable in these positions from the alkaline hydrolysate of bacteriophage; the amount of non-specific material present would have allowed the detection of RNA if present in an amount equivalent to more than 1% of the DNA content of the virus.

Chromatograms of the gipsy-moth virus hydrolysed whole, without isolation of its nucleic acid, showed no uracil, thus confirming the absence of RNA from this virus. A polynucleotide integrated from the found amounts of the DNA bases would be equivalent to 15·2% by weight of the virus, and account for 88% of the total virus phosphorus. Since similar purine-pyrimidine analyses of purified DNA's have usually succeeded in accounting for only about 90% of nucleic acid phosphorus (Wyatt, 1951*b*), it is probable that almost all of the virus phosphorus is included in its DNA. In one preparation, DNA was estimated by the diphenylamine reaction (Dische, 1930), which indicated the presence of 16·6% DNA after a downward correction of 4% had been applied to allow for the different purine/pyrimidine ratios in virus DNA and the thymus DNA standard (see Discussion).

DISCUSSION

Although the deoxypentose nucleic acids of only a small number of bacteria and viruses have been examined, these included organisms representative of several unrelated groups, and some generalizations are possible. Whereas DNA from a number of animals and at least one higher plant has a common pattern of composition, with more adenine and thymine than guanine and cytosine, a relatively small content of 5-methylcytosine, and a purine/pyrimidine ratio very close to unity (Wyatt, 1951*b*), this pattern is not shared by the nucleic acids of microbial origin. These are marked by the absence of 5-methylcytosine, by wide diversity in the relative proportions of the other bases, and by considerable

variation in the purine/pyrimidine ratio. There is some evidence that the composition of DNA varies less between related than between unrelated organisms, but many more nucleic acids will have to be analysed before we can state how closely these variations may parallel phylogenetic relationship.

Chargaff *et al.* (1950) have observed that DNA's from avian tubercle bacilli, *Serratia marcescens* and *Bacillus Schatz*, unlike those of higher animals, all contain an excess of guanine and cytosine over adenine and thymine, and on the strength of this they have proposed that DNA's may be divided into two main classes, the 'AT type' characteristic of animals and the 'GC type' of bacteria. According to this system, the bacteriophages would have the AT type and the gipsy-moth virus the GC type; *E. coli*, however, would have to be classified as 'transitional'. Again, only a more extensive survey of DNA composition can show whether this apparent grouping is more than fortuitous.

The variations in purine/pyrimidine ratio, whatever biological significance they may have, are of considerable practical importance, for the diphenylamine reaction commonly used for estimating DNA depends on the colour produced by purine-bound deoxypentose (Dische, 1930). If thymus nucleic acid, with a purine/pyrimidine ratio very close to 1·00, is used as the reference standard, estimations of DNA having a different purine/pyrimidine ratio will require correction by a corresponding factor. For example, estimations of DNA from phage T_3 will require downward correction by 3%—not, in this case, a very large error, but one which may possibly account for the recent finding by Cohen & Arbogast (1950) of enough DNA, estimated with diphenylamine, in phage T_3 to account for up to 105% of the total phosphorus of the virus. The error with phage T_4 or with *E. coli* would be greater still.

The compositions of DNA from host and bacteriophage are strikingly different and it is apparent that the bacterial nucleic acid could not be transferred directly in virus synthesis. This accords with the observations of Cohen (1948) and of Kozloff & Putnam (1950) that about 70–80% of the virus phosphorus is derived from inorganic phosphorus taken up from the medium after infection of the host. Kozloff & Putnam, however, consider that the remaining 20–30% of virus phosphorus comes from the bacterial DNA although this transfer may proceed via nucleotides.

SUMMARY

1. Deoxypentose nucleic acids have been isolated from the human and bovine strains of *Mycobacterium tuberculosis*, *Escherichia coli* (strain B/r), the bacteriophages T_3 and T_4 and the gipsy-moth polyhedral virus. Using paper chromatography the purine and pyrimidine compositions of these nucleic acids have been determined quantitatively.

2. Adenine, guanine, cytosine and thymine are present in all the nucleic acids, but 5-methylcytosine could not be detected.

3. The nucleic acids from different microorganisms show a wide range of nucleotide composition differing from the general structural pattern found among the deoxypentose nucleic acids of higher organisms.

4. The compositions of deoxypentose nucleic acids from the bacteriophages T_3 and T_4 are distinct, although similar, but differ greatly from that of the host (*Escherichia coli*).

We wish to express thanks to the Wellcome Research Laboratories for the gift of dried tubercle bacilli, and to Prof. J. Komárek, Charles University, Prague, who facilitated the collection of diseased gipsy-moth caterpillars as a source of polyhedra. One of us (G. R. W.) was engaged on this project while on transfer of research from the Division of Entomology, Department of Agriculture, Ottawa, Canada.

REFERENCES

Allen, R. J. L. (1940). *Biochem. J.* **34**, 858.
Bergold, G. (1947). *Z. Naturforsch.* **2b**, 122.
Bergold, G. (1950). *Viruses* 1950, ed. M. Delbruck, p. 60. California Institute Technology, Pasadena.
Bergold, G. & Pister, L. (1948). *Z. Naturforsch.* **3b**, 406.
Chargaff, E. (1950). *Experientia*, **6**, 201.
Chargaff, E., Zamenhof, S., Brawerman, S. & Kerin, L. (1950). *J. Amer. chem. Soc.* **72**, 3825.
Cohen, S. S. (1948). *J. biol. Chem.* **174**, 295.
Cohen, S. S. & Arbogast, R. (1950). *J. exp. Med.* **91**, 607.
Demerec, M. & Fano, U. (1945). *Genetics*, **30**, 119.
Dische, Z. (1930). *Mikrochemie*, **8**, 4.
Hoagland, C. L., Lavin, G. I., Smadel, J. E. & Rivers, T. M. (1940). *J. exp. Med.* **72**, 139.
Johnson, T. B. & Coghill, R. D. (1925). *J. Amer. chem. Soc.* **47**, 2838.
Kozloff, L. M. & Putnam, F. W. (1950). *J. biol. Chem.* **182**, 229.
Markham, R. (1942). *Biochem. J.* **36**, 790.
Markham, R. & Smith, J. D. (1949). *Biochem. J.* **45**, 294.
Mirsky, A. E. & Pollister, A. W. (1946). *J. gen. Physiol.* **30**, 117.
Putnam, F. W. (1950). *Science*, **111**, 481.
Putnam, F. W., Kozloff, L. M. & Neil, J. C. (1949). *J. biol. Chem.* **179**, 303.
Robinow, C. F. (1945). Addendum to *The Bacterial Cell*, p. 353. By Dubos, R. J. Cambridge, Mass.: Harvard University Press.
Schmidt, G. & Thannhauser, S. J. (1945). *J. biol. Chem.* **161**, 83.
Sevag, M. G., Lackman, D. B. & Smolens, J. (1938). *J. biol. Chem.* **124**, 425.
Vischer, E., Zamenhof, S. & Chargaff, E. (1949). *J. biol. Chem.* **177**, 429.
Witkin, E. M. (1946). *Proc. nat. Acad. Sci., Wash.*, **32**, 59.
Wyatt, G. R. (1950). *Nature, Lond.*, **166**, 237.
Wyatt, G. R. (1951*a*). *Biochem. J.* (in the Press).
Wyatt, G. R. (1951*b*). *Biochem. J.* (in the Press).

(Expt. Cell Res. Suppl. 2, 201 (1952))

SPECIFICITY IN THE COMPOSITION OF NUCLEIC ACIDS*

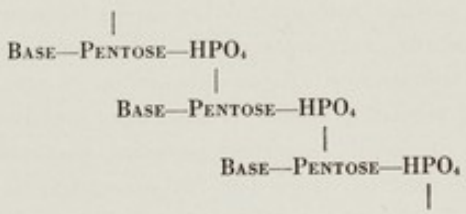
G. R. WYATT

Laboratory of Insect Pathology, Sault Ste. Marie, Ontario

NUCLEIC ACID STRUCTURE RECONSIDERED

Among the major chemical cell components, the nucleic acids remain today probably the least understood. This may seem surprising, since the basic units in the structure of nucleic acids are fewer in number than those of proteins, carbohydrates, or lipids, but this apparent simplicity was deceptive and led to a falsely simplified view of these substances which tended to inhibit enquiry for many years. Recently it has come to be realized that nucleic acids are very complex molecules possessing considerable biological specificity, the physico-chemical bases of which are currently being sought.

Below are listed the known constituent units of nucleic acids, with a diagrammatic indication of their linkage into polynucleotide chains:



	Ribonucleic acids (RNA)	Desoxyribonucleic acids (DNA)
Pentose	Ribose	Desoxyribose
Purines	Adenine	Adenine
	Guanine	Guanine
Pyrimidines	Cytosine	Cytosine
	Uracil	Thymine
		5-Methylcytosine

Most of these constituent substances have been recognized for many years. It seems likely that D-ribose and D-2-desoxyribose are the exclusive carbohydrate components of RNA and DNA respectively, for all nucleic acids give color reactions

*Contribution No. 26, Division of Forest Biology, Science Service, Department of Agriculture, Ottawa, Canada.

that partially characterize these sugars, and no evidence has yet been brought forward for the presence of any others. They have been unequivocally identified, however, in nucleic acids from only a few sources (14). The recent recognition of 5-methylcytosine as a new pyrimidine component of many desoxypentose nucleic acids (45) should serve as a warning against any complacent view that no new components remain to be discovered.

There are a number of possible bases for specificity in nucleic acids. The nature of the internucleotide linkages and the structure and configuration of polynucleotide chains are not yet definitely settled for any nucleic acid, so their possible contribution to specificity, great though it may be, cannot be usefully discussed at this stage. The order of sequence of different nucleotides in the chains is undoubtedly important and is being investigated, by analysis of products of partial hydrolysis, in several laboratories (28, 47). Our present knowledge, however, is only sufficient to emphasize the complexity of the problem. A simpler, if more superficial, approach to the differentiation of nucleic acids is by estimation of the relative amounts of the several purine and pyrimidine bases they contain. Results from this type of analysis, particularly with DNA, form the subject of this paper.

As soon as the major purine and pyrimidine bases from nucleic acids had been identified, early in this century, attempts were made to determine their quantitative relations. Chemical fractionation of these substances is not easy, as anyone will agree who has attempted to isolate them by chemical means, or who has analyzed commercial preparations of them. Nevertheless, some reasonably complete analyses were obtained, two of the best of which for "thymonucleic acid" (which may have come from thymus or fish sperm) are shown in Table 1. The proportions of the

TABLE 1
Composition of Thymonucleic Acid by Chemical Isolation of the Bases

Base	% Dry wt	Molar ratios*	% Dry wt	Molar ratios*
	Steudel (39)		Levene and Mandel (24)	
Adenine	10.68	1.30	8.17	0.98
Guanine	9.01	0.98	9.15	0.98
Cytosine	4.26	0.63	7.00	1.02
Thymine	8.33	1.09	8.00	1.02

*Molar ratios are calculated to total 4.00.

bases were not far from equimolecular. Similar results, with uracil instead of thymine, were obtained from nucleic acids of the pentose type from plants (wheat germ (33) and yeast (22)). As the researchers recognized the inaccuracy of the methods and were influenced by concepts borrowed from the chemistry of small molecules, they assumed that their results indicated equimolecular proportions of

four bases. This idea grew into the "tetranucleotide hypothesis" which postulated groups of two purine and two pyrimidine nucleotides, in fixed order, as the basic units of nucleic acid structure. It was implicit that there could be only one RNA and one DNA. This simplified view is dogmatically stated in Levene and Bass' monograph (23), and dominated thought on nucleic acids for many years.

More recently, physical measurements have shown that nucleic acids can have very high molecular weights—of the order of 10^6 in the case of thymus DNA. At the same time, biological evidence has accumulated suggesting that they play roles in heredity and protein synthesis which seem to require some specificity of structure. These developments engendered a new critical attitude toward their chemistry, and the tetranucleotide theory was questioned, in the first instance by Gulland (16). Evidence obtained by various methods during the past few years has all run counter to the old theory.

The road to more accurate knowledge of the composition of nucleic acids was finally opened by the development of two modern techniques, partition chromatography and ultraviolet spectrophotometry. These now provide simple and precise micro-methods for the analysis of both types of nucleic acid. The feasibility of separating nucleic acid derivatives by chromatography on filter paper was first demonstrated by Vischer and Chargaff (40) and by Hotchkiss (18), and during the past three years there has followed the publication of quantitative methods and results from a number of laboratories. All investigators agree that the composition of nucleic acids is not consistent with the tetranucleotide hypothesis, and most have recognized quantitative differences between nucleic acids from different sources.

QUANTITATIVE ANALYSIS OF NUCLEIC ACIDS

This symposium is not the place for detailed descriptions of technique, but a few remarks on the merits of various procedures may be of general interest. It has been found that truly quantitative separations of nucleic acid derivatives can be obtained by chromatography on filter paper, and for routine analytical work this method has the advantages of simplicity and sensitivity over the use of columns. Ion-exchange columns have proved better for special problems and preparative purposes. For detection of the spots on paper chromatograms, one can utilize either their fluorescence (17) or their absorption in the ultraviolet (26); the latter technique has been found very convenient, since it is sensitive to less than 1 μg of purine base and provides a permanent photographic record of each chromatogram.

As a solvent for the separation of the free bases on paper chromatograms, aqueous isopropanol (65% in 2.0 *N* HCl) (46) is very satisfactory, since the two purines and four pyrimidines commonly found in nucleic acids can all be resolved in a single one-dimensional run. The strong acid is helpful by increasing the solubility of guanine, of which only small amounts will migrate quantitatively in neutral or

weakly acidic or basic solvents. This mixture also resolves the nucleosides and some, but not all, of the nucleotides.

The nucleic acid derivatives eluted from chromatograms are estimated by their ultraviolet absorption in the quartz spectrophotometer. The method is sensitive and precise, but depends on accurate knowledge of extinction coefficients, and there is some disagreement in the literature on these (18, 34, 36, 41, 46). Agreement on these coefficients, re-determined with substances of established purity, would permit better comparison of results from different laboratories.

The problem of obtaining quantitative hydrolysis of nucleic acids has been troublesome, and many methods have been used. The purine bases are easily liberated from both RNA and DNA with dilute mineral acid, but splitting of the pyrimidine nucleosides requires much more vigorous treatment which frequently leads to deaminated or otherwise degraded products (36, 42). Separate hydrolyses for the purines and the pyrimidines have often been used, but since several methods are now known which permit estimation of all the bases from a single hydrolysate, only these will be discussed.

From RNA, free purines along with pyrimidine nucleotides can be obtained by hydrolysis for 1 hr in *N* HCl at 100° C (36). This method is not applicable to DNA, but the latter can be split completely to free bases by pure formic acid at 175° C for 30 min (46). A treatment yielding free bases from both types of nucleic acid has been described by Marshak and Vogel (31), using 70% $HClO_4$ at 100° for 1 hr. This method gives consistent results, although the yields of the bases differ significantly from those obtained by hydrolysis with formic acid (Table 2), and this must be allowed for when results from the two methods are compared.

TABLE 2
Composition of Bovine DNA According to Various Authors

Author	Molar ratios (total = 4.00)					% of P accounted for
	A	T	G	C	MC	
Chargaff et al. (10)	1.20	1.09	0.96	0.74	—	90
Marshak & Vogel (31)	1.19	1.15	0.83	0.82	—	96
Daly, Allfrey, & Mirsky (13)	1.11	1.15	0.93	0.81	—	99
Hurst, Marko, & Butler (20)	1.08	1.14	0.905	0.835	0.046	96
Wyatt (HCOOH)	1.13	1.12	0.86	0.85	0.053	91
Wyatt ($HClO_4$)	1.14	1.04	0.92	0.84	0.05	99

While enzymic methods of hydrolysis have usually given rise to complex mixtures of products, improved knowledge of purified enzymes is increasing their usefulness, and a quantitative enzymic method is being used in Butler's laboratory (19, 20). A specific phosphodiesterase obtained from venom of Russell's viper enables re-

duction of nucleic acid to mononucleotides, which can be separated on anion exchange columns and estimated by means of their phosphorus content. Because of its freedom from strong chemical treatments and its dependence on phosphorus estimations rather than ultraviolet absorption measurements, this method provides a useful check on the other techniques.

By suitable micro-techniques in conjunction with paper chromatography, purines, pyrimidines, and phosphorus can be estimated from a sample containing less than 0.5 mg of nucleic acid. In the work to be described on insect viruses, the standard errors of the means of triplicate analyses, each using 2 to 3 mg of virus, containing 0.3 to 0.5 mg of DNA, have regularly been less than $\pm$ 1.5%. This is, of course, a measure of the reproducibility of the method, not of its absolute accuracy.

The proportions of the purines and pyrimidines in DNA from bovine tissues as estimated in several laboratories are listed in Table 2. In all cases, the ratios of the several bases have been calculated to an arbitrary total of 4.00; this procedure has no logical basis now that it is known that not all nucleic acids contain only four bases, but it will be used throughout this paper for sake of uniformity. Where DNA from more than one bovine tissue has been analyzed in one laboratory, mean results are given, since no significant differences have been noted among tissues of one species. The figure for percentage of total phosphorus accounted for, on the assumption of a 1:1 molar relationship between bases and phosphorus in the nucleic acid, gives some measure of the completeness of an analysis. It may be pointed out, however, that technical errors can as easily make this figure erroneously high as erroneously low, and that equimolarity of phosphorus and bases in all nucleic acids is not definitely established. Considering the diversity of the methods used, agreement between the results of different authors is fairly close.

Analyses of RNA's in several laboratories have shown that these too depart widely from tetranucleotide structure and contain different proportions of the purines and pyrimidines according to their source (9, 36, 42). Constitutional differences have even been found between RNA's of different strains of tobacco mosaic virus (27) and between those of nuclei and cytoplasm of one tissue (29).

PURINE AND PYRIMIDINE COMPOSITION OF DESOXYPENTOSE NUCLEIC ACIDS

I should now like to present some of my own results on the composition of nucleic acids of the desoxypentose type.

One of the first observations during these studies was the recognition of the pyrimidine 5-methylcytosine as a new component of certain DNA's (44, 46). This pyrimidine had been reported in 1925 by Johnson and Coghill (21) as a constituent of tuberculinic acid; hence, when in 1949 an extra substance was observed in DNA from herring sperm, 5-methylcytosine seemed a likely identity for it. On isolation,

the new substance proved indeed to have the properties of 5-methylcytosine and to be identical with the synthetic compound. It was present in nucleic acid from many sources, but not a trace could be found in tubercle bacilli. Its absence from tubercle bacillus DNA has also been determined by Vischer, Zamenhof and Chargaff (43) so Johnson and Coghill's original claim, which provided the first clue to its identity, was apparently erroneous.

Methylcytosine has been found in amounts much smaller than those of the other bases in the DNA of all higher animals and plants yet examined, but not in bacteria and viruses. A complete picture of its distribution, however, depends upon analysis of many more species from different phyla of organisms. The constancy of the amounts found in preparations from a given source, and the fact that these are unaltered by treatment of the nucleic acid with NaOH or ribonuclease, indicate that it is a true component of DNA. The corresponding nucleotide (12) and nucleoside (15) have now also been recognized.

The molar ratios of the bases in DNA from eight animal sources and wheat germ, estimated after hydrolysis with formic acid, are shown in Table 3. It is

TABLE 3
Composition of DNA's from Animal Sources and Wheat Germ.
Formic Acid Hydrolysis. (Wyatt (46))

Source	Molar ratios (total = 4.00)				
	A	T	G	C	MC
Calf thymus	1.13	1.11	0.85	0.85	0.052
Beef spleen	1.13	1.12	0.85	0.85	0.054
Bull sperm	1.15	1.09	0.89	0.83	0.052
Ram sperm	1.15	1.09	0.88	0.84	0.039
Rat bone marrow	1.15	1.14	0.86	0.82	0.044
Herring sperm	1.11	1.10	0.89	0.83	0.075
Locusta migratoria	1.17	1.17	0.82	0.83	0.008
Echinus esculentus sperm	1.24	1.18	0.78	0.74	0.071
Wheat germ	1.05	1.08	0.94	0.69	0.23

obvious that there is great similarity among the different preparations. The differences in proportions of the major bases between one nucleic acid and another are in some cases significant but in others, no greater than the experimental error. The figure for methylcytosine, however, is significantly different for every species, and is the same only in the three nucleic acids from bovine tissues, thymus, spleen, and sperm. This corroborates the conclusion of Chargaff (7), based on analyses in which methylcytosine was not included, that DNA composition is a species character, differing for different species of organisms, but not for different tissues of one species. It is also worth noting that there is greater similarity of DNA composition among the various vertebrates than there is between these and two invertebrates

examined, a locust and a sea urchin, and that the DNA of wheat differs strikingly from the others in its high content of methylcytosine. There is thus a suggestion that the degree of similarity in the DNA of various species may bear some relation to their phylogenetic proximity.

A large group of viruses pathogenic to insects provides an especially convenient subject for nucleic acid analysis. These viruses cause the diseases of caterpillars and sawfly larvae known as polyhedral diseases and capsule or granule diseases (3, 4, 5) because of the abundant crystalline inclusion bodies produced during infection. These inclusion bodies can be purified in relatively large quantities from the diseased insects. By dissolving them in very dilute alkali the viruses can be liberated and isolated in a state of purity surpassed probably only by the crystallizable viruses of plants.

A great many species of insects are subject to these diseases, over 100 having already been recorded (38), and in general each seems to have its own specific virus, so that the viruses are here designated by the names of their hosts. Those that have been analyzed contain about 15% of nucleic acid, all of the desoxypentose type (6, 37) and extremely little lipid or non-nucleic acid carbohydrate. They can satisfactorily be hydrolyzed whole with perchloric acid for purine-pyrimidine analysis, as the protein degradation products do not interfere with estimation of the bases.

The compositions of the nucleic acids of several of these viruses are shown in Table 4. Methylcytosine was not found. It is notable that there are very wide differences in nucleic acid composition between some species, while others are almost if not quite identical, and the identities are not found where they might be expected on biological grounds. Thus, there is no significant difference between

TABLE 4
Composition of DNA's of Insect Viruses.
Perchloric Acid Hydrolysis

Virus designation*	Number of analyses	Molar ratios (total = 4.00)			
		A	T	G	C
Pd	3	0.845 ± 0.009**	0.80 ± 0.009	1.225 ± 0.003	1.13 ± 0.005
Bm	3	1.17 ± 0.010	1.12 ± 0.013	0.90 ± 0.002	0.81 ± 0.005
Cpe	3	1.185 ± 0.013	1.105 ± 0.003	0.90 ± 0.008	0.81 ± 0.008
Ma	1	1.18	1.11	0.91	0.80
Cm	2	1.285	1.235	0.775	0.70
Ns	2	1.29	1.21	0.785	0.715

* Pd = Gypsy moth (*Porthetria dispar* L.) Polyhedral virus.
Bm = Silkworm (*Bombyx mori L.*) Polyhedral virus.
Cpe = Alfalfa butterfly (*Colias philodice eurytheme* Boisd.) Polyhedral virus.
Ma = Tent caterpillar (*Malacosoma americanum* Fabr.) Polyhedral virus.
Cm = European budworm (*Cacoecia murinana* Hb.) Capsule virus.
Ns = Pine sawfly (*Neodiprion sertifer* Geoff.). Polyhedral virus.
**Mean and its standard error.

the polyhedral viruses of the silkworm (a moth) and the alfalfa butterfly, insects which fall in different sub-orders, and a provisional analysis of tent caterpillar polyhedral virus suggests that it too has the same composition. On the other hand, the virus of gypsy moth differs very greatly from those of silkworm and tent caterpillar, insects falling in different families of the same sub-order. Again, there is great similarity in nucleic acid between the capsule virus of the European budworm and the polyhedral virus of the pine sawfly; here not only are different orders of insects involved, but the viruses are characterized by the production of different types of inclusion body.

These comparisons tend to cast doubt on any idea of a direct parallelism between DNA composition and taxonomic relationships of organisms. The relationships of these viruses to one another, however, are not well understood, and it does not necessarily follow that they parallel those of their hosts, although among such highly adapted host-specific parasites this seems likely. Studies on the serological relationships of these viruses may help in understanding the significance of their nucleic acid composition.

Among all these animal, plant, and viral nucleic acids, however, certain regularities in the pattern of variation are evident. As first pointed out by Chargaff, Zamenhof, Brawerman, and Kerin (11), DNA's may be grouped by composition into two main types, an "AT type" having adenine and thymine the major bases, and a "GC type" in which guanine and cytosine preponderate. Animal DNA's were all of the AT type, while both types occurred among micro-organisms. Chargaff (8) has further observed a remarkable constancy in the ratios of adenine to thymine, guanine to cytosine, and total purines to total pyrimidines among DNA's from a number of sources. These generalizations are supported by the analyses described above.

The ratios between the amounts of different bases in animal and wheat germ DNA's are presented in Table 5. Cytosine and methylcytosine are grouped together because the most regular ratios are thus obtained, a tendency being noted for a greater methylcytosine content in a nucleic acid to be balanced by a reduced amount of cytosine. The ratios A:T, G:C+MC, and purines:pyrimidines are virtually constant, being in all cases nearly equal to unity. The ratio A+T:G+C+MC, on the other hand, gives a figure which, along with the methylcytosine content, distinguishes one nucleic acid from another.

The corresponding ratios for the insect virus nucleic acids have been computed in Table 6. The same regularities are here evident, even though both AT and GC types are included. It should be noted that these viruses were hydrolyzed with perchloric acid, whereas formic was used for the animal DNA's. When tested on the same preparation of nucleic acids perchloric acid gave a greater yield of guanine and a lower yield of thymine, each by about 6%, than formic acid (Table

TABLE 5
Molar Proportions in DNA's of Animals and Wheat.
Formic Acid Hydrolysis

Source	A / T	G / C + MC	Purines / pyrimidines	A + T / G + C + MC
Calf thymus	1.02	0.95	0.99	1.27
Beef spleen	1.01	0.94	0.98	1.28
Bull sperm	1.05	1.01	1.03	1.26
Ram sperm	1.05	1.00	1.03	1.27
Rat bone marrow	1.01	1.00	1.00	1.33
Herring sperm	1.01	0.98	1.00	1.23
Locust	1.00	0.98	0.99	1.41
Echinus sperm	1.05	0.96	1.01	1.52
Wheat germ	0.98	1.02	1.00	1.15

TABLE 6
Molar Proportions in DNA's of Insect Viruses.
Perchloric Acid Hydrolysis

Virus designation*	A / T	G / C	Purines / pyrimidines	A + T / G + C
Pd	1.06	1.08	1.07	0.70
Bm	1.04	1.11	1.07	1.34
Cpe	1.07	1.11	1.09	1.34
Ma	1.07	1.13	1.10	1.34
Cm	1.04	1.11	1.08	1.71
Ns	1.07	1.11	1.08	1.67

* See Table 4.

2). If allowance is made for this difference in method, the constant ratios found among the insect viruses become almost identical with those found for the animal nucleic acids, so that these relationships appear to be characteristic of a number of nucleic acids from very diverse sources. Whether the closeness of these ratios to unity actually implies equal numbers of nucleotides cannot be stated until the systematic errors of the methods are better known.

The compositions of DNA from two species of bacteria are shown in Table 7.

TABLE 7
Composition of DNA's of Two Bacteria (Smith & Wyatt (37))

Species	Hydrolysis	Molar ratios (total = 4.00)			
		A	T	G	C
M. tuberculosis	HCOOH	0.72	0.78	1.16	1.35
E. coli	$HClO_4$	0.92	1.10	0.96	1.03

In neither of these was methylcytosine found. The nucleic acids of tubercle bacilli and of *E. coli* are clearly different from each other and from any of the other nucleic acids analyzed. They still exhibit some regularities of composition similar to those observed above (Table 8), although the A:T, G:C, and purine:pyrimidine ratios are here lower than those of animal and insect virus DNA's. How much regularity there may be in these ratios among bacteria cannot be guessed from so small a sample.

TABLE 8
Molar Proportions in DNA's of Two Bacteria

Species	Hydrolysis	$\frac{A}{T}$	$\frac{G}{C}$	$\frac{\text{Purines}}{\text{pyrimidines}}$	$\frac{A+T}{G+C}$
M. tuberculosis	HCOOH	0.92	0.86	0.88	0.60
E. coli	$HClO_4$	0.84	0.93	0.88	1.01

The nucleic acids of bacteriophages are of particular interest in view of current biochemical studies on the multiplication of these viruses. Marshak (30) has recently reported the absence of cytosine from phage T_2. Smith and Wyatt (37), however, reported analyses of phages T_2 and T_5 which indicated that both contained appreciable amounts of cytosine, although it was much less than the other bases. In order to reconcile these results, a re-investigation of the nucleic acids of some bacteriophages is under way. It is possible that contamination with host DNA has caused trouble. It is at any rate clear that the DNA's of certain of these viruses are very differently composed from that of *E. coli*, a fact which precludes any direct transfer of nucleic acid from host to virus. The phage DNA's also do not seem to fit into the pattern of constant ratios described above.

A print of a chromatogram of the purines and pyrimidines of some typical DNA's and yeast RNA is shown in Plate I.

MOLECULAR NATURE OF DNA

Analytical results of the sort that has been presented naturally suggest attempts at calculation of the minimum molecular weight that can be fitted by the empirical ratios, and the numbers of nucleotides of each type in such a unit. The corresponding problem for amino-acid residues in proteins has been tackled with some success, but the difficulties are greater for nucleic acids than for proteins. This is because of the small number of different building units, and because the difficulty of interpreting physical measurements on such asymmetric particles as those of DNA makes the deduced molecular weights unreliable.

The majority of estimates from physical data indicate molecular weights of the order of 10^6 for DNA's. This corresponds to some 3000 nucleotides, or, if four kinds are present in roughly equal amounts, 700-800 of each. It is obvious that

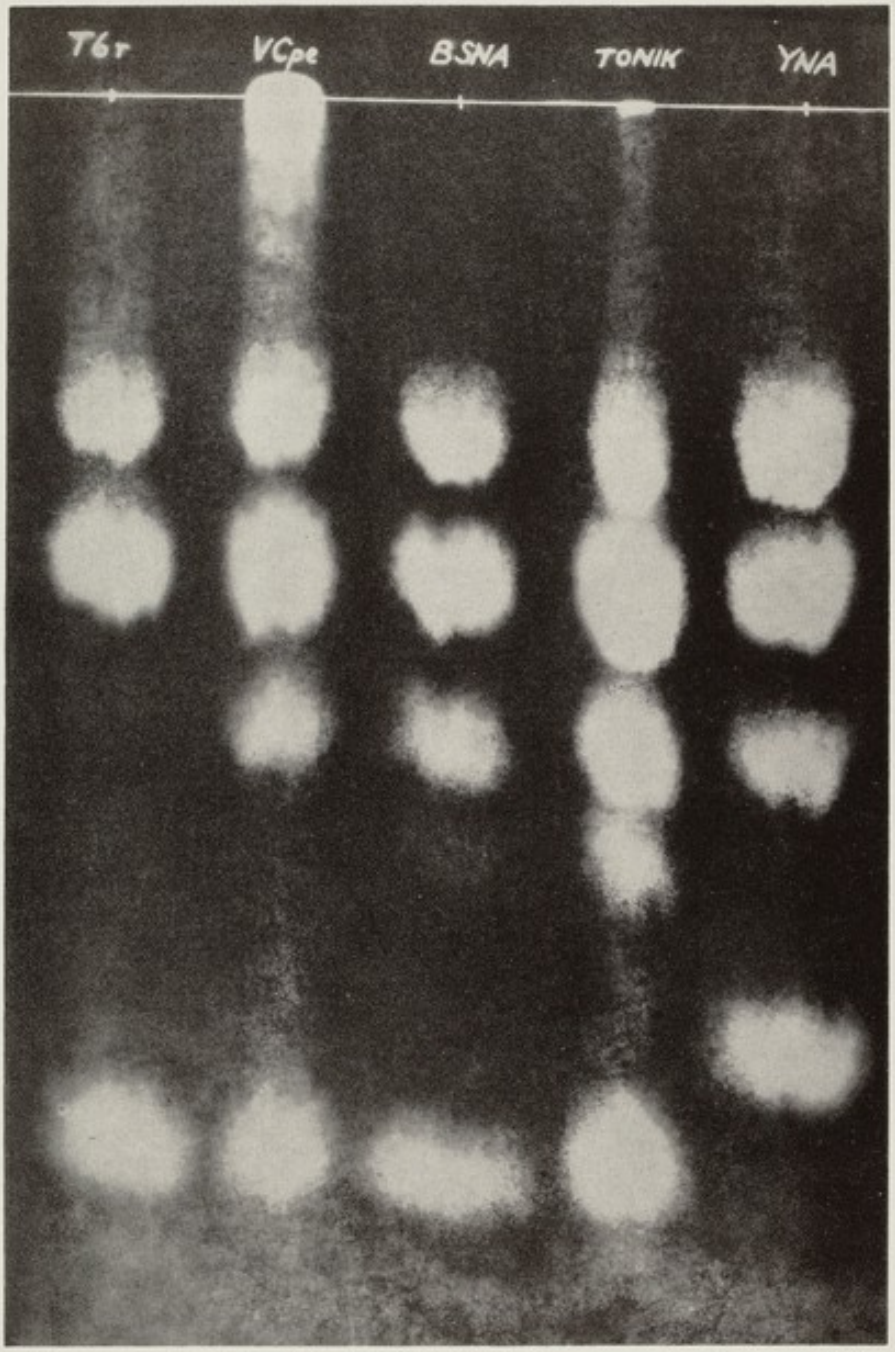


Plate I. A contact print, taken with ultraviolet light, of a paper chromatogram of the purines and pyrimidines of some typical nucleic acids. T_6r: *E. coli* bacteriophage T_6r; VCpe: alfalfa butterfly polyhedral virus; BSNA: beef spleen DNA; TONIK: wheat germ DNA; YNA: yeast RNA.
1: Guanine; 2: Adenine; 3: Cytosine; 4: 5-Methylcytosine; 5: Uracil; 6: Thymine.

a few of one kind could be replaced by another and escape detection by present methods of analysis. In such a nucleic acid, it is clearly futile to attempt to compute from the analysis a minimum molecular weight or the exact number of nucleotides of each kind in the molecule. The small content of methylcytosine in some nucleic acids, however, can provide a basis for calculation, for if it is assumed that the molecules are all of one type, there must be at least one methylcytosine nucleotide in each. Some relative nucleotide frequencies and corresponding minimum molecular weights on this basis are given in Table 9. The nucleotide numbers can be only

TABLE 9
Approximate Nucleotide Frequencies and Minimum Molecular Weights in Some DNA's, Based on the Content of Methylcytosine

Source	Relative numbers of nucleotides						Min. mol. wt.
	A	T	G	C	MC	Total	
Beef spleen	21	21	16	16	1	75	23,000
Herring sperm	15	15	12	11	1	54	17,000
Wheat germ*	9	9	8	6	2	34	11,500
Locust	146	146	102	104	1	499	155,000

* At least 2 methylcytosine nucleotides must be assumed, to obtain ratios which agree with the analysis within ± 5%.

approximate since allowance must be made for experimental error. The minimum molecular weights implied by the methylcytosine contents, although far below the molecular weights estimated for DNA, are sufficient to imply a remarkable specificity of structure. It is, indeed, difficult to imagine the synthesis of a polynucleotide chain having one residue of methylcytosine regularly in every 500, as implied by the composition of locust DNA.

An alternative hypothesis is that the molecules are not all of one type, but differ in their content of methylcytosine and presumably of the other bases also. There is no reliable evidence for the homogeneity of any DNA preparation; apparent centrifugal and electrophoretic monodispersity cannot be trusted because of boundary-sharpening effects. Singer (35) has found that a mixture of DNA's from *Escherichia coli* and bacteriophage T_2 appears homogeneous in the ultracentrifuge and during electrophoresis. In the case of RNA, there is good evidence that preparations are heterogeneous (1) and that the nucleic acid exists in different forms in the cell (2, 29). In this symposium, Dr. Bendich* has now put forward the first direct evidence of metabolic and structural heterogeneity of DNA from one tissue. It seems likely, therefore, that DNA preparations contain mixtures of molecular species, and the analyses which have been described represent merely the aggregate composition of these.

*See "Studies on the Metabolism of Nucleic Acids," p. 181.

COMPOSITION AND FUNCTION OF DNA

How can these observations on the composition of desoxypentose nucleic acids be related to present concepts regarding the function of these substances? It is generally considered that nucleic acids are essential for the multiplication of self-reproducing elements in the cell, RNA being involved in the production of cytoplasmic proteins, and DNA in the hereditary material of the nucleus. The lines of evidence which tend to link DNA with the genes have been reviewed in this symposium by Dr. Stern,* and may be summarized as:

(1) The localization of DNA in the chromosomes

(2) The metabolic stability of DNA as a cell component

(3) The production of mutations by ultraviolet irradiation, by X-rays, and by radiomimetic substances which also depolymerize DNA

(4) The production of "transformations" or mutations in certain bacteria by DNA preparations

(5) Constancy in the amount of DNA per set of chromosomes, as a species character.

To these may now be added, from the results from Chargaff's laboratory and those described in this paper:

(6) Constancy of DNA composition, as a species character.

By compiling data on the DNA content per nucleus of a large number of animal species, Mirsky and Ris (32) have demonstrated some correlation between these quantities and the evolutionary relationships of the species. The analytical results now suggest that there is also a general increase in complexity of DNA with evolutionary development of organisms, for it has been noted that the DNA's of certain bacterial viruses may be composed principally of three kinds of nucleotide, while those of other viruses and bacteria contain four, and those of some higher animals and plants have five. Many more species need to be examined, particularly from among the lower groups of animals and plants, to establish the validity of this generalization.

To carry speculation a step further, certain of these lines of evidence, in particular (3) and (4), are consistent with the idea that DNA may play a part in determining the special properties of individual genes. If this is so, each nucleus must contain a number of unique forms of DNA, for it is inconceivable that a single molecular type could possess the characteristics of all the genes of an organism. Each DNA preparation would then contain a number of molecular species perhaps even equal to the number of genes possessed by the source organism, and the analytical results would represent merely the mean composition of the total.

*See "Problems in Nuclear Chemistry and Biology," p. 1.

There is one feature of the results of purine-pyrimidine analyses that is consistent with the idea that we are dealing with the mean composition of numerous gene-specific nucleic acids. This is the observation of much greater constancy of composition among higher organisms than lower ones. It is notable, for example, that there is more similarity in DNA composition between such distantly related metazoa as locust, sea urchin, and ox than there is within a group of closely related lower organisms such as the polyhedral viruses of insects. This could follow from the greater tendency to average out among the large number of genes of higher organisms than among the relatively few genes of viruses. If, for example, the T-even bacteriophages possess 15 to 30 genetic units, as Luria and Dulbecco (25) have postulated, a single mutation might conceivably be reflected in a detectable change in the total nucleic acid composition, whereas among the thousands of genes of higher organisms a great many mutations would be required to alter the mean composition measurably.

It is difficult, however, to reconcile the complexity implied by this concept of nucleic acids with the apparent regularity connoted by the curiously constant ratios between certain of the bases. If the nucleic acids of individual genes do differ in composition, their variation must follow a fixed pattern by which these ratios are kept unchanged. We can scarcely even speculate upon how this could occur.

Also difficult to interpret is the apparently identical DNA composition of certain biologically distinct insect viruses. However, equality of composition of nucleic acids does not necessarily imply biological identity. Other possible determinants of specificity about which we know next to nothing are the order of sequence of nucleotides and the native configuration of the polynucleotide chains. The concept of gene-specific DNA's, then, can at present be put forward only as a tentative hypothesis, to stand or fall with evidence from other directions.

These analyses of the purine and pyrimidine composition of nucleic acids represent one approach, for which methods are now available, to a vast field which may prove to be the meeting point of chemistry and genetics.

References

1. Bacher, J. E. and Allen, F. W., *J. Biol. Chem.* **184**, 511 (1950)
2. Barnum, C. P. and Huseby, R. A., *Arch. Biochem.* **29**, 7 (1950)
3. Bergold, G., *Z. Naturforsch.* **2b**, 122 (1947)
4. Bergold, G., *Z. Naturforsch.* **3b**, 338 (1948)
5. Bergold, G. H., "Insect Viruses" *in* Viruses 1950. Pasadena, California, 1950
6. Bergold, G., and Pister, L., *Z. Naturforsch.* **3b**, 406 (1948)
7. Chargaff, E., *Experientia*, **6**, 201 (1950)
8. Chargaff, E., *Federation Proc.* **10**, 654 (1951)
9. Chargaff, E., Magasanik, B., Vischer, E., Green, C., Doniger, R. and Elson, D., *J. Biol. Chem.* **186**, 51 (1950)
10. Chargaff, E., Vischer, E., Doniger, R., Green, C. and Misani, F., *J. Biol Chem.* **177**, 405 (1949)

11. Chargaff, E., Zamenhof, S., Brawerman, G. and Kerin, L., *J. Am. Chem. Soc.* **72**, 3825 (1950)
12. Cohn, W. E., *J. Am. Chem. Soc.* **72**, 2811 (1950)
13. Daly, M. M., Allfrey, V. G. and Mirsky, A. E., *J. Gen. Physiol.* **33**, 497 (1950)
14. Davidson, J. N., The Biochemistry of the Nucleic Acids. Methuen, London, 1950
15. Dekker, C. A., and Elmore, D. T., *J. Chem. Soc.* 2864 (1951)
16. Gulland, J. M., Barker, G. R. and Jordan, D. O., *Ann. Rev. Biochem.* **14**, 175 (1945)
17. Holiday, E. R. and Johnson, E. A., *Nature* **163**, 216 (1949)
18. Hotchkiss, R. D., *J. Biol. Chem.* **175**, 315 (1948)
19. Hurst, R. O., Little, J. A. and Butler, G. C., *J. Biol. Chem.* **188**, 705 (1951)
20. Hurst, R. O., Marko, A. M. and Butler, G. C., personal communication, 1951
21. Johnson, T. B. and Coghill, R. D., *J. Am. Chem. Soc.* **47**, 2838 (1925)
22. Levene, P. A., *Biochem. Z.* **17**, 120 (1909)
23. Levene, P. A. and Bass, L. W., Nucleic Acids. Chem. Catalog Co., New York, 1931
24. Levene, P. A. and Mandel, J. A., *Biochem. Z.* **10**, 215 (1908)
25. Luria, S. E. and Dulbecco, R., *Genetics* **34**, 93 (1949)
26. Markham, R. and Smith, J. D., *Biochem. J.* **45**, 294 (1949)
27. Markham, R. and Smith, J. D., *Biochem. J.* **46**, 513 (1950)
28. Markham, R. and Smith, J. D., *Research* **4**, 344 (1951)
29. Marshak, A., *J. Biol. Chem.* **189**, 597 (1951)
30. Marshak, A., *Proc. Natl. Acad. Sci. U. S.* **37**, 299 (1951)
31. Marshak, A. and Vogel, H. J., *J. Biol. Chem.* **189**, 597 (1951)
32. Mirsky, A. E. and Ris, H., *J. Gen. Physiol.*, **34**, 451 (1951)
33. Osborne, T. B. and Harris, I. F., *Z. physiol. Chem.* **36**, 85 (1902)
34. Ploeser, J. McT. and Loring, H. S., *J. Biol. Chem.* **178**, 431 (1949)
35. Singer, J. S., personal communication, 1951
36. Smith, J. D. and Markham, R., *Biochem. J.* **46**, 509 (1950)
37. Smith, J. D. and Wyatt, G. R., *Biochem. J.* **49**, 144 (1951)
38. Steinhaus, E. A., Principles of Insect Pathology. McGraw-Hill, New York, 1949
39. Steudel, H., *Z. physiol. Chem.* **49**, 406 (1906)
40. Vischer, E. and Chargaff, E., *J. Biol. Chem.* **168**, 781 (1947)
41. Vischer, E. and Chargaff, E., *J. Biol. Chem.* **176**, 703 (1948)
42. Vischer, E. and Chargaff, E., *J. Biol. Chem.* **176**, 715 (1948)
43. Vischer, E., Zamenhof, S. and Chargaff, E., *J. Biol. Chem.* **177**, 429 (1949)
44. Wyatt, G. R., *Nature* **166**, 237 (1950)
45. Wyatt, G. R., *Biochem. J.* **48**, 581 (1951)
46. Wyatt, G. R., *Biochem. J.* **48**, 584 (1951)
47. Zamenhof, S. and Chargaff, E., *J. Biol. Chem.* **187**, 1 (1950)

Discussion

Oster: The idea that there are different compositions of nucleotides in nucleic acid may not be consistent with the X-ray results obtained by Dr. D. P. Riley and myself at the Royal Institution, London. We examined the X-ray "powder" diagram of highly polymerized desoxyribose nucleic acid. Nucleic acid dried over P_2O_5 shows only three broad diffraction rings in the neighborhood of 3-5 Å. A slight increase in moisture content makes no change but at 44% (by volume) of nucleic acid, there suddenly appears a magnificent display of some twenty sharp lines corresponding to spacings between about 3 and 16 Å. The spacing of 3.4 Å observed by

Astbury and Bell (*Nature* **141**, 747 (1938)) is, according to our data, actually a whole series of lines which they failed to resolve. These spacings do not change appreciably with increased water content indicating that they correspond to spacings within the molecules themselves. At higher moisture content there appear two larger spacings which vary with water content corresponding to inter-particle distances (RILEY, D. P. and OSTER, G., *Biochim. Biophys. Acta* **7**, 526 (1951)).

The sharpness of the diffraction pattern shows that the nucleic acid (with sufficient amount of water) is a substance of extreme regularity. I feel that if the regularity occurred only in the backbone of the highly polymerized chain molecule and the side groups (the purines, pyrimidines and sugar) were placed randomly along the chain, such a large number of spacings as observed would not be present. In any case, the excellent analyses carried out by Dr. Wyatt will be extremely valuable when attempts are made to try to analyze the meaning of the spacings.

Wyatt: I should like to ask Dr. Oster what types of groupings the X-ray method is capable of detecting. Can it differentiate between the nucleotides containing different bases? If not, you would get the same diffraction pattern even if you had different arrangements of nucleotides along the chain.

Oster: No. At this early stage of the analysis of the X-ray diagrams we cannot distinguish between nucleotides containing different bases. If, however, one could incorporate say, iodine, which scatters X-rays about one hundred times more than does carbon or nitrogen, into one of the bases without destroying the general arrangement of the nucleotides then one should be able, by means of X-ray diffraction, to locate the spacings between such bases. If no diffraction was forthcoming from such a procedure then I would agree that the bases were not arranged in a regular pattern along the chain.

Wyatt: I don't know enough about the nature of the chemical bonds involved to comment very much upon this but I think it is quite possible, particularly if you have a spiral structure, that the purine and pyrimidine residues are sticking out free so that they don't interfere with each other. Then you could have a regular spacing down the backbone of the chain, in spite of the differences in sequence.

Stern: The molecular models which I mentioned this morning confirm what Dr. Wyatt just said. If you construct such a model you will find that the bond angles and the distances of different groupings along the backbone are more or less independent of the sequence of the individual nucleotides. So, I agree with the inference that even if the sequence were changed, purely physical methods, like X-ray diffraction, would hardly be able to pick that up. I would like to mention that one of the principles of variation which I mentioned this morning is actually open to experimental proof: that the different DNA's in the same chromosome might differ simply by chain length. This is a problem which the polymer chemists have studied in great detail. We have very sensitive criteria for chain length measurements and for polydispersity measurements by flow birefringence for instance. A very simple experiment which I would suggest would be to isolate, by exactly the same methods, DNA from two living forms, one a very low form, where one would expect very few genes to be present, and one a highly developed form. The distribution curves for both DNA preparations could then be determined with regard to molecular weights and chain lengths. If this idea is correct there should be a positive correlation between these physical parameters and the number of different genes present in the chromosomes of the two species.

Haurowitz: If the nucleic acids consist of a chain of alternate molecules of carbohydrate and phosphoric acid, in which each carbohydrate molecule is linked to a purine or pyrimidine base, then the distribution of electric charges will be the same for all nucleic acids, and it is hard to imagine any specificity of such molecules. If, however, loops can be formed by hydrogen bonds between the phosphate groups and basic groups of the purine or pyrimidine bases, then an infinite number of specific nucleic acids is imaginable. The molecular models mentioned by Dr. K. G. Stern may be helpful in investigating this type of specificity.

Stern: There is one point which we mentioned in our earlier paper (*Yale J. Biol. and Med.* **19**, 937 (1947)) along these lines. That is, if you look at the formula of the typical tetranucleotide model you will find that the two purine bases and the two pyrimidine bases, both in DNA and in PNA, show a complementary structure at one part of the ring which makes for hydrogen bond formation. One has the CO group while the other has an NH_2 group. Now, if they vary in the sequence of purine-pyrimidine then these two complementary groups within the same polynucleotide chain are so far away from each other, a distance of over 7 Å according to Astbury's measurements, that hydrogen bonding is not very probable. But, if you rearrange them so that you have the sequence purine-pyrimidine and that the nucleotide side chains are projecting from the same side of the backbone, then these complementary groups approach each other over a distance of 3.5 Å where hydrogen bonding is just still possible. I might say that D. O. Jordan (A.C.S. meeting, San Francisco, 1949) has discussed the possibility of hydrogen bonding particularly with regard to adjacent polynucleotide chains. He feels that the fact that he gets different electrometric titration curves when he titrates forward and then titrates backwards, is due to the breakage of hydrogen bonds during titration. I think that there is a good probability of both types of bonding (i.e., of intra- and inter-molecular hydrogen bonds) in nucleic acids.

Bendich: There is recent evidence obtained by Dr. Cavalieri in our laboratory indicating that there are covalent linkages present in the nucleic acid structure, which occur between the phosphate and amino and hydroxyl groups of the bases. In addition, there is evidence for salt-type linkages, i.e., between phosphate and amino groups (Cavalieri, L. F. and Angelos, A., *J. Am. Chem. Soc.*, **72**, 4686 (1950); Cavalieri, L. F., Kerr, S. E. and Angelos, A., ibid, **73**, 2567 (1951)).

[From THE BIOCHEMICAL JOURNAL, Vol. 48, No. 5, pp. 581–584, 1951]

PRINTED IN GREAT BRITAIN

Recognition and Estimation of 5-Methylcytosine in Nucleic Acids

By G. R. WYATT*

Agricultural Research Council Plant Virus Research Unit, Molteno Institute, University of Cambridge

(*Received* 19 *June* 1950)

The pyrimidine 5-methylcytosine was first reported as a constituent of a nucleic acid by Johnson & Coghill (1925), who prepared from hydrolysed tuberculinic acid a crystalline picrate which was identified as 5-methylcytosine picrate by microscopical examination. However, when Vischer, Zamenhof & Chargaff (1949) examined nucleic acid from avian tubercle bacilli by chromatography, they could not detect this substance. Nevertheless, Hotchkiss (1948) noted on paper chromatograms of hydrolysed thymus nucleic acid a small spot whose ultraviolet spectrum and chromatographic behaviour led him to suggest that it might be 5-methylcytosine.

As already reported in a preliminary note (Wyatt, 1950), I have found by a simple chromatographic method that 5-methylcytosine seems to occur in constant amounts in deoxypentose nucleic acids (DNA) from animals and from at least one higher plant, but has not so far been found in DNA from microbial sources. The isolation, partial characterization and estimation of this substance will now be described. Quantitative results will be presented along with the method and results of estimating the other pyrimidine components of some DNA's in the succeeding paper (Wyatt, 1951).

EXPERIMENTAL

Isolation of the new substance

The purines and pyrimidines set free by hydrolysis of DNA with formic acid (175° for 30 min.) were separated on paper chromatograms run with aqueous 65% (v/v) *iso*propanol 2N in respect of HCl as described by Wyatt (1951), and their positions detected by the photographic technique of Markham & Smith (1949). In preliminary experiments with DNA from ox spleen and from herring sperm, it was noted that in addition to guanine, adenine, cytosine, thymine and sometimes a trace of uracil, another substance formed a faint spot which moved slightly more rapidly on the chromatograms than cytosine. Tests had shown that the substance was not cytosine deoxyriboside, since its concentration was not appreciably diminished by increasing the hydrolysis time, and was not an aromatic amino-acid, since it gave no colour with ninhydrin. The isolation of a sample was undertaken as follows.

40 g. of crude herring-sperm DNA were freed of purines by methanolysis with HCl (Levene & Bass, 1931, p. 110). The filtrate from the precipitated purine hydrochlorides was evaporated to dryness, dissolved in 6N-HCl, and heated at 125° for 2 hr. to complete liberation of the pyrimidines. The hydrolysate was clarified with activated charcoal, evaporated to dryness under diminished pressure, redissolved in water and again evaporated. The tarry residue was taken up in water and freed of H_3PO_4 by neutralization with $Ba(OH)_2$ and filtration. After evaporation to a small volume, a crystalline precipitate separated out, consisting mainly of thymine; after further reduction in volume a precipitate containing cytosine, thymine and some of the new substance formed. The filtrate was made alkaline with NaOH and the remaining pyrimidines were precipitated with $AgNO_3$. After removal of Ag^+ with HCl this fraction was found to contain the new substance along with some thymine, uracil, adenine and much cytosine.

This concentrate was fractionated by paper chromatography. The solution was applied as a band across the top of sheets of Whatman no. 3 filter paper, which were run as chromatograms in *iso*propanol-aqueous HCl. In order to shorten exposure time, because of the thickness of Whatman no. 3 paper, Ilford Document Paper, no. 60, was used for printing these chromatograms instead of Reflex Document Paper, no. 50. As the chromatograms were overloaded (about 10 mg. of each substance can be separated on an 18 × 22 in. sheet of Whatman no. 3 paper), resolution was poor, but the bands containing the new substance were eluted and the eluate chromatographed in *n*-butanol-aqueous NH_3. The eluate this time contained only one ultraviolet-absorbing substance, and on evaporation deposited microscopic crystals.

Comparison with 5-methylcytosine

On paper chromatograms the new substance moved slightly more rapidly than cytosine in all the solvents tested, which included *iso*propanol-aqueous HCl, *n*-butanol-water, *n*-butanol-aqueous formic acid or NH_3, and *iso*amyl alcohol-aqueous NH_3. The position relative to cytosine was similar regardless of pH, but the separation from cytosine increased the higher the alcohol used. This strongly suggested a compound bearing ionizable groupings identical with those of cytosine, but differing by an additional alkyl group, and 5-methylcytosine seemed a likely possibility.

An authentic sample of 5-methylcytosine was not at the time available, but some synthetic 2-amino-4-hydroxy-5-methylpyrimidine (5-methylisocytosine) was kindly given by Prof. A. R. Todd, F.R.S. This had absorption spectra different from those of the unknown, and on chromatograms moved slightly faster.

If the unknown were 5-methylcytosine, it would be converted to thymine by deamination with HNO_2. Accordingly, 0·5 mg. of the substance in 1 ml. water was mixed with 2M-$Ba(NO_2)_2$ (2 ml.) and glacial acetic acid (0·5 ml.). After 6 hr. at room temperature, the excess acetic acid was removed by evaporation, and Ba^{++} with H_2SO_4. The filtrate from $BaSO_4$ was evaporated to dryness and dissolved in 0·1 ml.

* Present address: Laboratory of Insect Pathology, Box 490, Sault Ste Marie, Ontario, Canada.

of water, and spots were applied to paper for chromatography. A portion of the synthetic 5-methylisocytosine was treated in the same way. Both were found to have been converted quantitatively to a substance identical with thymine in its chromatographic movement (Fig. 1) and ultraviolet absorption spectrum.

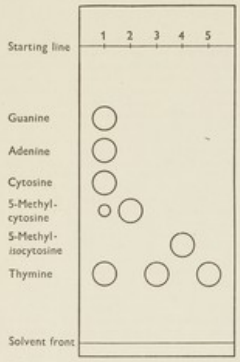


Fig. 1. Diagram of a paper chromatogram run in *iso*propanol-aqueous 2 N-HCl. Solutions applied at positions 1–5 were: (1) hydrolysate of herring-sperm deoxypentose nucleic acid (DNA); (2) supposed 5-methylcytosine isolated from DNA; (3) product of deamination of (2); (4) synthetic 5-methylisocytosine; (5) product of deamination of (4).

There are only three substances which would yield thymine on deamination: 5-methylcytosine, 5-methylisocytosine and 2:4-diamino-5-methylpyrimidine. The unknown had been found different from the second of these, and was most unlikely to be the third, since the possession of an additional amino group would cause slower movement in acid and more rapid movement in ammoniacal solvents on the chromatograms. This made it reasonably certain that the new substance was 5-methylcytosine.

After these experiments had been completed, a sample of 5-methylcytosine synthesized by the method of Wheeler & Johnson (1904) was obtained through the kindness of Dr A. Pircio, Fordham University, New York. This proved to be identical with the natural substance in its movement in both acid and ammoniacal chromatographic solvents and in its ultraviolet absorption spectra in neutral, acid and alkaline solution (Fig. 2). These spectra are also identical with those recently published by Hitchings, Elion, Falco & Russell (1949) for 5-methylcytosine prepared by a new synthesis.

Since the absorption curves and chromatographic behaviour provide evidence for the identity of the natural and synthetic substances, isolation of a sample sufficient for accurate elementary analysis has not been attempted. Owing to the low capacity of paper chromatograms, only a small portion of the concentrate obtained by precipitation had been fractionated by chromatography. Separation on starch columns was also tried, but their capacity proved

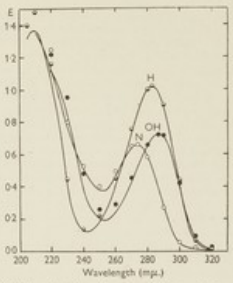


Fig. 2. Ultraviolet absorption spectra of 5-methylcytosine in 0·1 N-HCl (▭), in 0·1 N-NaOH (●) and at pH 7·0 (○). The curves represent the absorption of synthetic 5-methylcytosine (0·013 mg./ml.), and the points that of natural 5-methylcytosine eluted from chromatogram spots and read against paper blanks.

little greater than that of filter paper. Fractional crystallization proved useless for separation of cytosine and 5-methylcytosine. If it were desired to isolate a greater amount of natural 5-methylcytosine, a better starting material would be DNA from wheat germ, which contains one-third as much 5-methylcytosine as cytosine. The final fractionation might be accomplished by counter-current distribution or on ion-exchange columns.

Ultraviolet absorption characteristics of 5-methylcytosine

For determination of its ultraviolet absorption characteristics, some synthetic 5-methylcytosine was purified through the picrate, recrystallized from methanol and from water, and dried at 110° for weighing. Standard solutions were made up in water (taken to pH 7·0 with a little NaOH), 0·1 N-HCl and 0·1 N-NaOH. Molecular extinction coefficients at maxima and minima are given in Table 1. In acid solution there is an additional maximum at about 211 mμ. The characteristic shifts in acid and alkali correspond closely to those of cytosine, the wavelengths of absorption maxima being in each case longer by 6–8 mμ., and those of the minima by 3–4 mμ. The shift from the 'acid' to the 'neutral' spectrum takes place at a pH just below 7, and that from

the 'neutral' to the 'alkaline' at about pH 12, so that the curve for pH 7 (Fig. 2) and that for pH 11 (Hitchings *et al.* 1949) are almost identical.

Table 1. *Ultraviolet absorption characteristics of 5-methylcytosine*

(ε is molecular extinction coefficient.)

	Maxima		Minima	
Solvent	Wavelength (mμ.)	ε	Wavelength (mμ.)	ε
Water	274	6200	252	3680
0·1 N-HCl	283	9810	242	950
0·1 N-NaOH	287	6870	253	1670

Method of estimation

The amount of methylcytosine in animal nucleic acids is so small that estimation by elution of the spot on chromatograms of unfractionated hydrolysate is inaccurate, especially if traces of absorbing impurities are present. A procedure has therefore been used which is in effect two-dimensional chromatography, the substances being concentrated after running in the first dimension. Hydrolysate of 6–8 mg. of DNA is applied in a band 25 cm. long across the top of a sheet of paper, which is then run in isopropanol-aqueous-HCl, dried and printed. A transverse band containing all of both cytosine and methylcytosine is cut from the chromatogram; one end of this band is cut to a point, and the other end is placed in a trough containing water in a chromatogram tank. The pyrimidines move with the water front and are eluted quantitatively in the first 0·5 ml. to drip from the bottom of the paper strip, which is collected in a small tube. This is then evaporated to dryness and redissolved in 0·1 N-HCl (0·04 ml.). Measured volumes are placed as spots on filter paper and run in *n*-butanol-aqueous NH_4, in which cytosine and methylcytosine are well separated from each other and from impurities extracted from the paper of the first chromatogram. The separated spots are eluted and their pyrimidine contents estimated on the Beckman spectrophotometer; the amount of cytosine serves as a standard for referring the amount of methylcytosine to the whole nucleic acid. In this way methylcytosine can be estimated when it comprises as little as 0·1% of the original nucleic acid.

Distribution of 5-methylcytosine

The various animal DNA's so far analysed (including mammals, a fish, an insect and an echinoderm) all contain methylcytosine, in amounts characteristic of the species source, and varying from 0·008 to 0·075 mol./4 mol. of nucleotide. The only DNA from a higher plant analysed, that of wheat germ, contains much more, 0·23 mol./4 mol. of nucleotide (Wyatt, 1951). None could be detected in DNA from the following microbial sources: human and bovine tubercle bacilli, *Escherichia coli*, *Esch. coli* bacteriophages T_2 and T_4, and the virus of the polyhedral disease of gipsy moth larvae; none or extremely little is present in DNA from baker's yeast. Nor could any be found in the ribonucleic acids of yeast and of turnip yellow-mosaic virus.

DISCUSSION

In the present state of knowledge as to the structure and function of nucleic acids nothing can be said as to the possible function of 5-methylcytosine. The amounts in which it occurs, however, varying with the source but constant from a given source, suggest that it is an essential constituent of certain DNA's, and no accident of enzyme action. The demonstration that these nucleic acids do contain this pyrimidine in addition to their well known components extends the evidence of recently published analyses that their structure is a matter of considerable complexity and worthy of much further investigation. In particular, it would be interesting to know whether the observation that methylcytosine is present in higher organisms but lacking in micro-organisms holds generally, and if so, at what point in the evolutionary scale it first appears.

It is ironical that the nucleic acid of tubercle bacilli, from which Johnson & Coghill (1925) originally reported 5-methylcytosine, should be one of those which do not contain it. As a clue to the source of their error, it is interesting to note that Vischer *et al.* (1949) examined by chromatography a specimen of methylcytosine obtained from Prof. T. B. Johnson and found it to contain 42% of cytosine, so that the crystalline '5-methylcytosine picrate' with which the picrate from tuberculinic acid was compared may well have been chiefly that of cytosine.

SUMMARY

1. An aminopyrimidine which on deamination gives rise to thymine has been discovered as a new constituent of some deoxypentose nucleic acids (DNA). The ultraviolet absorption spectra and chromatographic behaviour of this substance are identical with those of synthetic 5-methylcytosine, and the tentative conclusion is drawn that it is the latter.

2. By a simple chromatographic technique, this pyrimidine can be estimated when occurring in amounts as small as 0·1% of a nucleic acid.

3. Ultraviolet absorption spectra of 5-methylcytosine have been determined.

4. The new pyrimidine occurs in amounts characteristic of the source in all of the animal DNA's and the one plant DNA analysed, but is apparently lacking from DNA of bacteria and viruses, and from ribonucleic acids.

I wish to thank Dr J. D. Smith for information on his analyses of nucleic acid from *Esch. coli* and its bacteriophages, Dr R. Markham for helpful advice and discussion, and Dr K. M. Smith, F.R.S., for providing facilities for the work. The work was carried out while I was on transfer of research from the Forest Insect Investigations Unit, Division of Entomology, Department of Agriculture, Canada, to the Agricultural Research Council Plant Virus Research Unit, Molteno Institute, Cambridge. This paper is Contribution No. 2701, Division of Entomology, Science Service, Department of Agriculture, Ottawa, Canada.

REFERENCES

Hitchings, G. H., Elion, G. B., Falco, E. A. & Russell, P. B. (1949). *J. biol. Chem.* **177**, 357.
Hotchkiss, R. D. (1948). *J. biol. Chem.* **175**, 315.
Johnson, T. B. & Coghill, R. D. (1925). *J. Amer. chem. Soc.* **47**, 2838.
Levene, P. A. & Bass, L. W. (1931). *Nucleic Acids.* New York: Chemical Catalog Co., Inc.
Markham, R. & Smith, J. D. (1949). *Biochem. J.* **45**, 294.
Vischer, E., Zamenhof, S. & Chargaff, E. (1949). *J. biol. Chem.* **177**, 429.
Wheeler, H. L. & Johnson, T. B. (1904). *Amer. chem. J.* **31**, 591.
Wyatt, G. R. (1950). *Nature, Lond.*, **166**, 237.
Wyatt, G. R. (1951). *Biochem. J.* **48**, 584.

contained most vitamin A in Exp. 1, in which the lowest amounts of vitamin A were administered. In Exps. 2, 3, 4 and 5 the mean concentrations of vitamin A in the kidneys only varied between 9·9 and 17·1 i.u./g., as compared with 49–3340 i.u. for the mean concentrations in the liver. Previous work has shown, however, that with doses of vitamin A approaching the toxic level much higher concentrations in the kidneys are readily attained (Davies & Moore, 1934).

REFERENCES

Davies, A. W. & Moore, T. (1934). *Biochem. J.* **28**, 288.
Johnson, R. M. & Baumann, C. A. (1947*a*). *Arch. Biochem.* **14**, 361.
Johnson, R. M. & Baumann, C. A. (1947*b*). *Fed. Proc.* **6**, 265.

The Quantitative Analysis of Deoxypentose Nucleic Acids. By G. R. Wyatt (introduced by R. Markham). (*A.R.C. Plant Virus Research Unit, Molteno Institute, Cambridge*)

A method simpler than that of Vischer & Chargaff (1948) has been developed for the quantitative analysis by paper chromatography of the purine and pyrimidine bases in deoxypentose nucleic acids. The bases freed by hydrolysis in formic acid at 175° for 30 min. are separated on one-dimensional chromatograms, using as the solvent an aqueous solution containing *iso*-propanol 65% (v/v) and HCl 2·0N. The spots are detected by the photographic method of Markham & Smith (1949), eluted, and estimated spectrophotometrically.

In addition to adenine, guanine, thymine, and cytosine, certain nucleic acids contain 5-methyl-cytosine, which has been isolated and identified by deamination to thymine and by comparison with synthetic 5-methyl-cytosine. It is present in all the animal deoxyribonucleic acids yet analysed (from mammals, fish, and insects) in small amounts characteristic of the source, and in nucleic acid from wheat germ, but it has not been detected in microbial nucleic acids. In agreement with Vischer, Zamenhof & Chargaff (1949), and contrary to Johnson & Coghill (1925), none could be found in nucleic acid from tubercle bacilli.

The following are preliminary results:

Molar proportions of the bases in deoxypentose nucleic acids*

Source	Adenine	Guanine	Thymine	Cytosine	5-Methyl-cytosine
Calf thymus	1·10	0·86	1·15	0·89	0·052
Beef spleen	1·12±0·006†	0·86±0·006	1·16±0·008	0·86±0·011	0·052±0·002
Ram sperm	1·14	0·89	1·11	0·86	0·042
Herring sperm	1·10±0·015	0·91±0·009	1·16±0·021	0·86±0·006	0·073±0·002
Locusts (whole)	1·15	0·82	1·18	0·86	<0·01
Wheat germ ('Bemax')	1·10	1·00	1·16	0·74	0·23
Tubercle bacilli (human and bovine)	0·70	1·16	0·78	1·36	0·00
Escherichia coli	0·92	0·81	1·20	1·06	0·00
E. coli bacteriophage T_5	1·30	0·83	1·45	0·42	0·00
Virus of gypsy moth polyhedral disease	0·86	1·20	0·92	1·02	0·00

* Calculated to an arbitrary total of 4·00 (excluding methyl-cytosine).
† Standard error of the mean of six or seven independent sets of estimations.

The similarity in composition of the animal deoxyribonucleic acids is very striking, although the differences between those from different species are statistically significant. Microbial nucleic acids show much greater divergences. The possibility must be considered that the preparations may not be homogeneous, but attempts at fractionation have not yet succeeded.

REFERENCES

Johnson, T. B. & Coghill, R. D. (1925). *J. Amer. chem. Soc.* **47**, 2838.
Markham, R. & Smith, J. D. (1949). *Biochem. J.* **45**, 294.
Vischer, E. & Chargaff, E. (1948). *J. biol. Chem.* **176**, 715.
Vischer, E., Zamenhof, S. & Chargaff, E. (1949). *J. biol. Chem.* **177**, 429.

The Metabolism of ^{15}N-labelled Urea in the Cat. By R. E. Davies and H. L. Kornberg. (*Medical Research Council Unit for Research in Cell Metabolism, Department of Biochemistry, The University, Sheffield* 10)

Until recently urea has been regarded as solely an end-product of mammalian nitrogen metabolism. Bloch (1946) tested this point by feeding ^{15}N-labelled urea to rats. He obtained a slight incorporation of ^{15}N in the proteins of the liver and the intestine but regarded this as the result of bacterial decomposition of urea in the gut. On the other hand, Leifer, Roth & Hempelmann (1948), who injected ^{14}C labelled urea into mice, recovered 20·8 % of the injected ^{14}C in the exhaled CO_2 and were thus led to assume that urea is not a stable end-product in the mammalian body. This has been confirmed by Skipper (1949).

Table 1 gives some of the results. The isotope content of all other tissues was between 0·003 and 0·008 atom % excess ^{15}N and accounted for 10 % of the injected ^{15}N. The amide-N fraction of the liver was significantly enriched and 2·3 % of the injected ^{15}N appeared in the urine in a form other than urea or ammonia.

These results show that at least 2·5 % of the urea N was metabolized. This figure is a minimum; Berenbom & White (1950) have shown that if ammonium glutamate, ^{15}N-labelled in the ammonium group, is injected intravenously into rats, most of the isotope is rapidly excreted as urea and that there is only slight incorporation into liver and kidney proteins. Furthermore, since any ammonia arising from the hydrolysis of urea will be largely reconverted to urea in the liver, the amount of isotope found in the non-urea non-ammonia fraction of the urine can represent only a minimum figure for the metabolism. The present results are therefore not incompatible with the value of 20·8 % breakdown obtained by Leifer *et al.* (1948), which also represents a minimum. The low isotope content of gastric mucosa suggests that the urea-urease system of this tissue plays no part in local synthetic mechanisms.

Table 1. *Distribution of* ^{15}N *after injection of* 215 *mg.* ^{15}N*-labelled urea* (32 *atom* % *excess*) *into a cat*

Fraction	N content (g./100 g.)	Total N (g.)	^{15}N content (atom % excess)	Percentage of injected ^{15}N	Method of estimation
Urine total N	0·94	2·35	1·075	81	Kjeldahl N
Urine urea N*	0·73	1·83	1·287	75	(*a*) Fosse (1916), Kiech & Luck (1928); (*b*) Van Slyke & Cullen (1916)
Urine ammonia N	0·15	0·13	0·042	0·17	Van Slyke & Cullen (1916)
Urine non-urea non-ammonia N	0·21	0·53	0·136	2·3	Kjeldahl N on urea- and ammonia-free urine
Liver total N	2·84	1·66	0·008	0·42	Kjeldahl N
Liver amide N	0·12	0·08	0·013	0·04	Schoenheimer, Ratner & Rittenberg (1939)
Gastric mucosa total N	2·04	0·15	0·003	0·01	Kjeldahl N

Recovery of injected ^{15}N: 92 %.

* Estimated by two different methods. Accuracy of mass spectrometer: ±0·003.

The following experiments were undertaken with reference to the problem of the role of gastric urease. ^{15}N-labelled urea, synthesized according to Cavalieri, Blair & Brown (1948), was injected into a cat weighing 2·0 kg., kept on a low-protein diet. The cat was chosen because the urease activity of ground gastric mucosa is much higher in cats than in rats (Luck, 1925; FitzGerald, 1946). The dose was 98·8 mg. urea on the first day and 116·4 mg. on the second. The urine was collected in conc. HCl under toluene. On the third day the cat was killed; the tissues were rapidly removed and stored at −12°. The ^{15}N content of various fractions of urine, blood, brain, diaphragm, heart, intestinal tract, kidneys, skin and hair, spleen and stomach was measured by mass spectrometry.

We are greatly indebted to Dr R. Bentley, who carried out the mass-spectrophotometric analyses.

REFERENCES

Berenbom, M. & White, J. (1950). *J. biol. Chem.* **182**, 5.
Bloch, K. (1946). *J. biol. Chem.* **165**, 469.
Cavalieri, L. F., Blair, V. E. & Brown, G. B. (1948). *J. Amer. chem. Soc.* **70**, 1240.
FitzGerald, O. (1946). *Nature, Lond.*, **158**, 305.
Fosse, R. (1916). *Ann. Chim. Phys.* **6**, 13.
Kiech, V. C. & Luck, J. M. (1928). *J. biol. Chem.* **77**, 723.
Leifer, E., Roth, L. J. & Hempelmann, L. H. (1948). *Science*, **108**, 748.
Luck, J. M. (1925). *Biochem. J.* **18**, 825.
Schoenheimer, R., Ratner, S. & Rittenberg, D. (1939). *J. biol. Chem.* **127**, 333.
Skipper, H. E. (1949). Private communication.
Van Slyke, D. D. & Cullen, G. E. (1916). *J. biol. Chem.* **24**, 117.

[From THE BIOCHEMICAL JOURNAL, Vol. 48, No. 5, pp. 584–590, 1951]
 PRINTED IN GREAT BRITAIN

The Purine and Pyrimidine Composition of Deoxypentose Nucleic Acids

By G. R. WYATT*
Agricultural Research Council Plant Virus Research Unit, Molteno Institute, University of Cambridge

(*Received* 19 *June* 1950)

Although the composition of nucleic acids has been subject to frequent investigation since these compounds were first isolated by Miescher in the last century, it is only through the application of techniques developed in the past few years that their accurate quantitative analysis has become possible. Early analyses by macrochemical methods (Steudel, 1906; Levene & Mandel, 1908) indicated that nucleic acids contained two purine and two pyrimidine bases in approximately equimolecular proportions. This gave rise to the tetranucleotide hypothesis, upheld by Levene & Bass (1931) and widely accepted, according to which the unit of nucleic acid structure is a group of four nucleotides comprising one of each of the four bases. Recent micro-analyses, however, by such methods as filter-paper chromatography and ultraviolet spectrophotometry, have shown that the composition of nucleic acids is not consistent with any such simplified theory, a conclusion which is in line with biological evidence of their intimate function in cell physiology and of the specific activities of certain of them, e.g. in inducing mutations in bacteria (Avery, MacLeod & McCarty, 1944; Boivin, 1947).

The feasibility of separating nucleic acid derivatives by paper chromatography was demonstrated by Vischer & Chargaff (1948*a*) and by Hotchkiss (1948). A method for the quantitative analysis of ribonucleic acids (RNA) was developed by Vischer & Chargaff (1948*b*), and later applied to deoxypentose nucleic acids (DNA) (Chargaff, Vischer, Doniger, Green & Misani, 1949). This involves separation of the purines and the pyrimidines: the former are precipitated on methanolysis with dry hydrogen chloride, and liberation of the latter is completed by formic acid at 175°. The two groups are then estimated on separate chromatograms.

A simpler procedure for the analysis of RNA has been described by Smith & Markham (1950), who use a short hydrolysis in N-hydrochloric acid to liberate purines along with pyrimidine nucleotides, all of which are then separated on one-dimensional paper chromatograms. This method is not applicable to DNA, from which nucleotides or nucleosides cannot be obtained quantitatively by chemical hydrolysis because of the lability of the deoxy-sugar. In the method now described, DNA is hydrolysed by formic acid to purines and pyrimidines, which are separated on single paper strips. Preliminary results from this method have already been reported (Wyatt, 1950), and the identification and estimation of the pyrimidine 5-methylcytosine in DNA is described in the preceding paper (Wyatt, 1951).

Since this work was done, Daly, Allfrey & Mirsky (1950) have published analyses of DNA from a number of sources, based on separations obtained with starch columns. Their results differ from those now presented in that they failed to recognize 5-methylcytosine, and inclined to the view that the differences between various nucleic acids were not significant.

METHODS

Detection of the spots of chromatograms

For detection of the spots of purines and pyrimidines on paper chromatograms the photographic technique of Markham & Smith (1949) was used, which consists in making contact prints of the dried chromatograms on reflex document paper with filtered ultraviolet light of wavelengths 254 and 265 mμ.

Chromatographic solvents

Although purines and pyrimidines can be separated by various neutral, ammoniacal or weakly acid solvents (a number are listed by Markham & Smith, 1949), such solvents

* Present address: Laboratory of Insect Pathology, Box 490, Sault Ste Marie, Ontario, Canada.

are not satisfactory for quantitative resolution of nucleic acid hydrolysates in amounts adequate for reliable estimation, because guanine 'tails' badly, owing to its low solubility. Many such solvents were tested. However, when Smith & Markham (1950) had shown that solvents containing relatively high concentrations of HCl could be used successfully on filter-paper chromatograms, systems of this type were tried. Increasing acidity has the effect of slowing the movement of all the substances and must be balanced by increasing the water content. The purines are held back more than the pyrimidines, and in solvents containing HCl stronger than about 1 N, adenine moves more slowly than cytosine, reversing the positions they take in neutral systems. After many mixtures had been tested, that selected for resolution of the bases from DNA was an aqueous solution containing *iso*propanol 65% (v/v) and HCl 2·0 N in the whole volume. This is made up as follows: to 65 ml. of absolute *iso*propanol (or 68 ml. of 95% *iso*propanol) at 16° add sufficient titrated conc. HCl to give 0·2 g.mol. of HCl, and water to 100 ml. The proportions are rather critical, and some empirical adjustment may be necessary; at higher temperatures the volume of *iso*propanol is reduced slightly. To obtain the best spacing of the spots the solvent should be used within a few days of mixing and the tank must be very well saturated with its vapour.

The R_F values of the common purines and pyrimidines in this solvent are given in Table 1, along with those of some nucleosides and nucleotides, to illustrate the effect on chromatographic movement of linkage to ribose and deoxyribose and of phosphorylation. These must be regarded merely as a guide to the relative positions of the spots, as absolute R_F's vary considerably.

Table 1. *R_F values in filter-paper chromatography of nucleic acid derivatives in* iso*propanol* (65 %, v/v)-*water*-HCl (2·0 N)

	R_F (Whatman no. 1 filter paper)
Purines:	
Guanine	0·25
Xanthine	0·25
Hypoxanthine	0·31
Adenine	0·36
Pyrimidines:	
Cytosine	0·47
5-Methylcytosine	0·55
Uracil	0·68
Thymine	0·77
Nucleosides and nucleotides:	
Cytosine riboside	0·50
Cytosine deoxyriboside	0·63
Cytidylic acid	0·61
Uridylic acid	0·79

Up to 75 μg. of each substance can be resolved from mixtures in 35 cm. movement of the solvent front. An incidental advantage of this solvent is that xanthine and hypoxanthine are estimated along with guanine and adenine respectively, so that the error caused by any deamination of the purines will be minimized. Before printing, the chromatograms must be dried until most of the HCl has evaporated. This can be done overnight by hanging them at room temperature, or for about 30 min. in a current of warm air, but excessive heat causes charring and must be avoided.

Estimation of the purines and pyrimidines

The procedure for estimation of the substances resolved on paper chromatograms is as follows. Triplicate spots of hydrolysate are applied to Whatman no. 1 filter paper from a micropipette (that used throughout these experiments contained 18·0 μl.); the chromatograms are run by the descending method until the front has moved about 35 cm. and are then dried and printed photographically. The positions of the spots are traced from the print on to the filter paper, and disks including them are cut out, along with blanks of equal size cut from a free 'lane' opposite the substances to be estimated; each is eluted in a test tube containing 5 ml. of 0·1 N-HCl. After standing for several hours, the eluates are read at their absorption peaks against the corresponding blanks in the Beckman spectrophotometer.

When tested with solutions of pure substances, this method has given recoveries of 96–100%. The range of variation of replicate spots on one chromatogram and of subsequent chromatograms of one material is generally less than 5%; occasional greater discrepancies have been traced to remnants of ultraviolet-absorbing detergent or other dirt on the test tubes used for elution. A smaller error may result from the absorption of the silica Beckman cells themselves, which varies from cell to cell (cf. Morton, 1949). The set of cells used in this work have an absorption peak at 243 mμ. with an optical density averaging 0·014 when the three more opaque cells are read against the most transparent. Their absorption difference at 250 mμ. is sufficient to cause an error of up to 3% in estimations of guanine; this has been corrected for in results here presented.

A technique for the estimation of minor components, such as methylcytosine, by elution from one chromatogram and running again with another solvent, has already been described (Wyatt, 1951).

Table 2. *Ultraviolet absorption data on purines and pyrimidines in* 0·1 N-*hydrochloric acid*

(mε is millimolecular extinction coefficient.)

Substance	Wave-length (mμ.)	mε	$E^{0.001\%}_{1\,cm.}$	$E^{0.001\%}_{1\,cm.}$ (Hotchkiss, 1948)
Adenine	260	13·0	0·96	1·01
Guanine	250	11·0	0·73	0·665
Uracil	260	7·9	0·705	0·72
Cytosine	275	10·5	0·95	0·91
Thymine	265	7·95	0·63	0·61
5-Methylcytosine	283	9·8	0·785	—

For use in estimation of the purines and pyrimidines separated by chromatography, extinction coefficients have been determined from pure samples. The sources of the material used for this are: guanine and uracil Hoffmann-La Roche; adenine, British Drug Houses Ltd.; thymine, commercial sample recrystallized; cytosine, isolated from herring-sperm nucleic acid, purified through the picrate and recrystallized; 5-methylcytosine, the gift of Dr A. Pircio, University of Chicago, purified through the picrate and recrystallized. Each gave only one spot on chromatography except cytosine, in which a trace of methylcytosine persisted; this, however, was too small to cause significant error. The samples were dried, weighed and dissolved in 0·1 N-HCl for extinction measurement. Following Hotchkiss (1948), readings were taken at the nearest multiple of 5 mμ. to the wavelength of maximum absorption, except for methylcytosine. Results are given in Table 2, along with values calculated from the data of Hotchkiss for comparison; all except guanine agree within 5% with Hotchkiss's values.

Quantitative hydrolysis of deoxypentose nucleic acid

Hydrolysis with formic acid. The value of pure formic acid in the hydrolysis of nucleic acids was shown by Visscher & Chargaff (1948*b*), who found that it liberated pyrimidines with much less deamination than was caused by strong mineral acids. For RNA, however, 2 hr. at 175° were required for complete splitting of the pyrimidine ribosides, and in this time significant destruction of the purines had occurred. This led to their using separate hydrolyses for the purines and the pyrimidines.

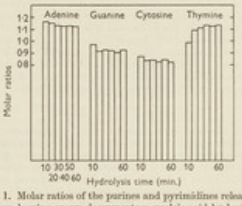


Fig. 1. Molar ratios of the purines and pyrimidines released from herring-sperm deoxypentose nucleic acid by hydrolysis in 98% formic acid at 175° for various lengths of time. For each hydrolysis time the yield of each base in molar units is expressed as a proportion of a total of 4·00.

It was known, however, that acid hydrolysis of DNA proceeds more rapidly than that of RNA, and in the hope of obtaining quantitative yields of all the bases from a single hydrolysis, shorter treatments with formic acid were tried. The purines are released from DNA by exceedingly short acid treatment, and of the pyrimidines cytosine is freed more quickly than thymine. At 165° in 98% formic acid recovery of cytosine reaches a maximum after 15 min., but the yield of thymine increases up to 2 hr. The results of a series of hydrolyses of herring-sperm DNA at 175° are shown in Fig. 1. Maximal yield of thymine is reached by 30 min., and there is no significant drop in recovery of any of the bases up to 60 min. (the higher initial ratios of adenine, guanine and cytosine are due to deficiency of thymine, since the four are calculated to constant totals). The experiment was repeated with a commercial thymus nucleic acid, with the same outcome. 30 min. at 175° in 98% formic acid therefore appears to give maximal yields of both purines and pyrimidines from DNA. As a further check on possible destruction of purines, the yield of adenine from DNA by this treatment was compared with that from 15 min. in aqueous N-HCl at 100°, and found to be the same.

To carry out this hydrolysis conveniently 1–30 mg. of DNA is weighed into a Pyrex tube of 13 mm. diameter, or a solution containing this amount is evaporated to dryness in the tube. Then 0·5 ml. of 98% formic acid is added and the tube is sealed off and heated in an electric muffle at 175° for 30 min. After cooling, the tube is opened cautiously in view of its contents being under pressure and the hydrolysate evaporated to dryness in the same tube. Aqueous N-HCl is then added in sufficient quantity to make a solution 3·4% (w/v) with respect to the original nucleic acid, allowance being made for the volume of the nucleic acid on the assumption that its specific volume is 0·6. The hydrolysate is allowed to dissolve and portions are applied to the chromatogram paper with a micropipette. It is important to use strong enough acid for dissolving the hydrolysate, as otherwise part of the guanine may remain out of solution.

Hydrolysis with perchloric acid. Since most of this work was done, Marshak & Vogel (1950) have reported quantitative hydrolysis of both RNA and DNA by 1·2 N-$HClO_4$ at 100° for 1 hr. This method has been tried. In the experiment summarized in Table 3, equal quantities of ox-spleen DNA were hydrolysed with formic acid as described above and with 72% (w/w) $HClO_4$ (11·1 N) at 100° for 1 hr., and analysed by chromatography simultaneously. The differences were small, and the total recovery of bases was nearly the same by each method. By repeating the comparative experiment, it was confirmed that $HClO_4$ gives consistently a slightly higher molar ratio of guanine and a lower one of thymine than formic acid; the reasons for this are not clear, but slight destruction of guanine by treatment with formic acid is probable.

Hydrolysis with $HClO_4$ has also been tried on RNA. A sample of the purified yeast nucleic acid which had been analysed by Smith & Markham (1950) was hydrolysed in

Table 3. *Recovery of bases from ox-spleen deoxypentose nucleic acid after hydrolysis with formic and with perchloric acids*

	Mol./atom P in hydrolysis with			Molar ratios* in hydrolysis with	
	HCOOH (*A*)	$HClO_4$ (*B*)	*A*/*B*	HCOOH	$HClO_4$
Adenine	0·257	0·259	0·99	1·13	1·12
Guanine	0·191	0·207	0·92	0·84	0·99
Cytosine	0·198	0·195	1·02	0·87	0·85
Thymine	0·254	0·250	1·02	1·12	1·08
Total	0·900	0·911	0·99	3·96	3·95

* Calculated to total 3·95, making 4·00 when 0·05 is added for 5-methylcytosine (cf. Table 8).

this way; the molar ratios found are compared in Table 4 with ratios calculated from Smith & Markham's (1950) published values by applying certain corrections recently determined by Dr R. Markham (personal communication). Considering that the two methods involve different hydrolyses, different chromatographic solvents, and the use of independently determined extinction values, the closeness of agreement is remarkable.

Table 4. *Composition of yeast ribonucleic acid hydrolysates*

	Molar ratios found by Smith & Markham (1950) (corrected—see Text)	Molar ratios found after hydrolysis with $HClO_4$ (means of 3 estimations)
Adenine	1·02±0·01*	1·04±0·01*
Guanine	1·19±0·01	1·22±0·005
Cytosine	0·83±0·02	0·82±0·01
Uracil	0·96±0·04	0·92±0·015

* Standard error of the mean.

Hydrolysis with $HClO_4$ is conveniently carried out as follows. To the air-dry nucleic acid in a small Pyrex tube is added a measured volume of 72% $HClO_4$ such that the concentration of nucleic acid in solution is 6–8%. The tube is closed with a glass stopper bound in place and is heated in a boiling-water bath for 1 hr. After cooling, the hydrolysate is diluted with an equal volume of water, and the tube is briefly centrifuged to sediment the residue of carbon. Samples can then be pipetted from the supernatant solution directly on to the chromatogram paper. In the *iso*propanol-aqueous HCl solvent, $HClO_4$ remains at the starting point and does not interfere with separation of the bases. Owing to the very hygroscopic nature of conc. $HClO_4$, it is advisable to determine the concentration of nucleic acid in the hydrolysate by estimation of P in samples taken at the same time as those taken for chromatography.

As a further check on possible destruction of the bases by either formic or perchloric acid, some experiments were performed to test the recovery of known amounts of pure bases added to nucleic acid and then subjected to hydrolysis. Table 5 shows the results of one such test. Portions of a solution containing adenine, guanine and thymine in 90% (w/w) formic acid (which is an excellent solvent for sparingly soluble bases such as guanine) were added to samples of ox-spleen DNA and evaporated to dryness, and the mixtures hydrolysed by both the formic acid and the $HClO_4$ techniques. Samples of the nucleic acid without the added bases were hydrolysed at the same time, and the quantities of the bases in the various hydrolysates and in the unhydrolysed solution of pure bases estimated by chromatography. P estimations on the hydrolysates gave a check on volume change by the $HClO_4$ solutions. The amounts recovered of the three added bases were all better than 95% by both methods of hydrolysis, so that little destruction during hydrolysis is indicated. It should be pointed out, however, that great accuracy cannot be expected from an experiment of this sort, which depends upon the comparison of three estimated quantities, each subject to experimental error, and the true yields may not be quite as nearly quantitative as these figures suggest. A previous similar experiment gave less satisfactory results, although recovery of each base was still better than 90%.

It is therefore concluded that essentially quantitative yields of both purines and pyrimidines from DNA can be obtained by treatment either with 98% formic acid at 175° for 30 min., or with 72% $HClO_4$ at 100° for 1 hr., and that the second method is suitable for RNA also. Since the latter method was not tried until its publication by Marshak & Vogel (1950), the former has been used in most of the analyses described in this paper. The use of $HClO_4$ has certain practical advantages, since the use of sealed tubes is not necessary, but difficulty has been experienced in hydrolysing impure nucleic acids with $HClO_4$, for the presence of protein interferes, retarding splitting of pyrimidine ribosides.

Table 5. *Recovery of added bases subjected to hydrolytic conditions in the presence of ox-spleen deoxypentose nucleic acid*

	μg./chromatogram spot			
	Adenine	Guanine	Cytosine	Thymine
Formic acid hydrolysis:				
Found in hydrolysed mixture	65·8	63·8	19·4	55·0
Accounted for by NA	30·7	25·4	19·3	28·4
Extra base found, by difference	35·1	38·4	—	26·6
Extra base added	36·6	39·2	0·0	26·3
Percentage recovery	96%	98%	—	101%
Perchloric acid hydrolysis:				
Found in hydrolysed mixture	67·2	67·2	19·4	53·1
Accounted for by NA	31·1	27·6	19·1	27·9
Extra base found, by difference	36·1	39·6	—	25·2
Extra base added	36·6	39·2	0·0	26·3
Percentage recovery	99%	101%	—	96%

Preparation of nucleic acids

For standardization of the analytical technique two preparations of DNA were made by methods avoiding the use of acid, alkali or heat. One was prepared from ox spleen (calf thymus not being available) by a method similar to that of Mirsky & Pollister (1946). The minced fresh spleen tissue was washed three times with cold 0·15 M-NaCl, then extracted with M-NaCl. Nucleoprotein was precipitated from the centrifuged extract by dilution with 6 vol. of water, then redissolved in M-NaCl and deproteinized by emulsification with $CHCl_3$-commercial octanol mixture (8:1, v/v) and centrifugation. After dialysis against water the nucleic acid was precipitated by addition of 1 vol. ethanol (90%, v/v) and washed successively with 90% (v/v) and absolute ethanol and ether. A sample for weighing was dried at 110° over P_2O_5 *in vacuo*.

DNA was prepared from herring testes by a similar procedure, except that papain was used for removal of the protamine, which fails to form a gel with $CHCl_3$. Some protein remained in this preparation.

The method of Mirsky & Pollister (1946) for preparation of crude nucleoprotein by extraction with M-NaCl and precipitation by dilution with water is applicable to most tissues (though not to mammalian sperm) and has been used in preparing most of the DNA preparations here studied. When the ratio of RNA to DNA in the original tissue is high, however, preparation of the latter in the pure state requires tedious reprecipitations accompanied by considerable loss. In making DNA from locusts, even when the minced tissue was washed four times with 0·15 M-NaCl and the nucleoprotein twice precipitated with water, some RNA remained.

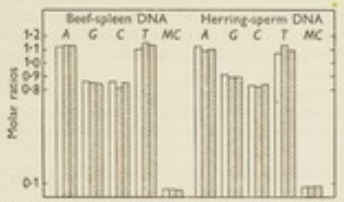


Fig. 2. Molar ratios of purines and pyrimidines from ox-spleen and herring-sperm deoxypentose nucleic acids, untreated (□), after treatment with N-NaOH (▨), and after treatment with ribonuclease (▩). *A*, adenine; *G*, guanine; *C*, cytosine; *T*, thymine; *MC*, 5-methylcytosine.

Since the fractionation of Schmidt & Thannhauser (1945) with N-NaOH has been widely used as an analytical procedure, it seemed worth while to discover whether this treatment might safely be applied to the preparation of DNA for analysis. Samples of the ox-spleen and herring-sperm DNA were dissolved in N-NaOH, left overnight at 37°, precipitated by acidification with glacial acetic acid and the addition of 1 vol. of ethanol, and analysed. In case this treatment might alter the composition by removal of traces of RNA, identical samples were incubated with ribonuclease and precipitated. The composition after each treatment is compared with the original composition in Fig. 2. The only significant change caused by the action of both alkali and ribonuclease on both nucleic acids is a slight increase in the ratio of thymine to the other bases, and this could easily result from the removal of a trace of contaminant RNA. The composition is otherwise not significantly altered, and the characteristic differences between the two nucleic acids are retained. Treatment with N-NaOH overnight at 37° has therefore been used wherever removal of RNA was necessary, and with preparations containing much RNA (such as nucleic acid from whole locusts) has been found much more effective than ribonuclease. Of the nucleic acids whose analysis is given below, NaOH was used in the preparation of those from bull sperm, ram sperm, rat-bone marrow, locusts and wheat germ.

The finding that the composition of DNA is unaltered by alkaline treatment does not, of course, diminish the importance of using mild methods of preparation when macromolecular properties are to be studied.

A sample of calf-thymus nucleic acid by a modification of the method of Hammarsten (1924) was the gift of Dr K. Bailey, DNA from rat-bone marrow was prepared by Dr C. Lutwak-Mann, and that from ram sperm in collaboration with Dr T. Mann, to all of whom thanks are due.

Elementary micro-analyses

Total N was estimated by the micro-Kjeldahl method, and P by the colorimetric methods of Allen (1940) and Tisdall (1922).

RESULTS

In Table 6 are shown the results of an analysis of calf-thymus nucleic acid with respect to purines and pyrimidines (after hydrolysis with formic acid), P and N. The P and N accounted for by the estimated purines and pyrimidines have been integrated, the former on the assumption of simple polynucleotide structure, and the calculated values expressed as percentages of total P and N.

The results of similar analyses of other nucleic acids with respect to P and N are summarized in Table 7.

Table 6. *Composition of calf-thymus nucleic acid* (Na *salt*)

		μg. mol./mg. dry wt.			
	% of dry wt.	Found	Calculated from bases	% of total accounted for	Molar* ratios
Adenine	10·00	0·741	—	—	1·11
Guanine	8·51	0·563	—	—	0·84
Cytosine	6·37	0·574	—	—	0·86
Thymine	9·46	0·751	—	—	1·13
Methylcytosine	0·46	0·037	—	—	0·056
Total bases	34·80	2·666	—	—	4·00
Phosphorus	9·4	3·03	2·67†	88	—
Nitrogen	14·5	10·36	9·86	95	—

* Calculated to an arbitrary total of 4·00.
† Assuming simple polynucleotide structure.

Table 7. *P and N contents of nucleic acid preparations with percentages of total* P *and* N *accounted for by estimated purines and pyrimidines*

Nucleic acid of	% of dry wt.		% of total accounted for by estimated bases (means of all experiments)	
	P	N	P	N
Calf thymus	9·4	14·5	90	94·5
Ox spleen	9·35	15·0	90	92·5
Ram sperm	7·4	15·9	88·5	68
Herring sperm	8·7	15·1	91	88
Wheat germ	9·7	14·9	84·5	92
Bull sperm	7·0*	—	90	—
Echinus sperm	8·6*	—	92	—

* Not specially dried.

The mean molar ratios of purines and pyrimidines in eight animal and one plant DNA are listed in Table 8, calculated in each case to an arbitrary total of 4·00. Where three or more independent analyses (separate hydrolyses and separate chromatograms) have been made, the standard error of the mean is shown.

Noting that Smith & Markham (1950) accounted for more than the total P of yeast RNA by their estimated bases, I have analysed a sample of their purified yeast nucleic acid by hydrolysis with perchloric acid, and although the ratios of the bases came very close to theirs (Table 4), the P accounted for by the bases was only 92 % of the total P. It seems most likely that both types of nucleic acid do contain bases and P in the 1:1 relationship usually assumed, and that the deficiency of estimated bases may be due to undiscovered error in the analytical techniques.

Significance of the molar ratios

The conclusion has been drawn by Chargaff, Zamenhof & Green (1950) that the composition of deoxyribonucleic acids varies in different species but not in different tissues of one species. The results of the present study tend to confirm this generalization and to strengthen it by the inclusion of 5-methylcytosine in the analyses.

Among the three bovine DNA preparations analysed those of thymus and spleen show no significant differences. The slightly increased purine/

Table 8. *Ratios of purines and pyrimidines from deoxypentose nucleic acids, after hydrolysis with formic acid*

Source of nucleic acid	No. of analyses	Molar ratios calculated to total 4·00				
		Adenine	Guanine	Cytosine	Thymine	Methylcytosine
Calf thymus	3	1·13±0·013	0·86±0·008	0·85±0·008	1·11±0·011	0·052±0·004
Ox spleen	7	1·13±0·006	0·85±0·006	0·84±0·011	1·13±0·008	0·054±0·002
Bull sperm	2	1·15	0·89	0·83	1·09	0·052
Ram sperm	3	1·15	0·88	0·84	1·09	0·039±0·001
Rat bone marrow	2	1·15	0·86	0·82	1·14	0·044
Herring sperm	6	1·11±0·015	0·89±0·009	0·83±0·006	1·10±0·021	0·075±0·002
Locusta migratoria (whole)	3	1·17±0·016	0·82±0·025	0·83±0·015	1·17±0·010	0·008±0·002
Echinus esculatus sperm	3	1·24±0·019	0·78±0·003	0·74±0·014	1·18±0·004	0·071
Wheat germ	2	1·06	0·94	0·69	1·08	0·23

DISCUSSION

The relation between bases and phosphorus

In all the nucleic acids analysed, more P is present than can be accounted for by assigning one phosphate radical to each molecule of estimated purine and pyrimidine. The fraction of the total P accounted for in this way averages 90 % (Table 7), and does not differ materially whether based on P estimated directly in samples of hydrolysate or on the P content of the original material. This discrepancy from 100 % seems greater than can be accounted for by destruction of the bases during hydrolysis, and suggests the presence of some phosphate additional to that in a simple polynucleotide structure. In an attempt to remove any loosely bound phosphate contaminant, a portion of ox-spleen DNA was reprecipitated ten times and analysed, but the fraction of total P accounted for by the bases was still only 91 %, so that little if any extraneous phosphate had been removed.

pyrimidine ratio in bull-sperm DNA as compared with the other two may be due to incipient autolysis, as the sperm was incubated with papain in an attempt to extract the nucleic acid before this was done with alkali; there is evidence that nucleases detach pyrimidine nucleotides preferentially (Overend & Webb, 1950). The methylcytosine ratio, however, is identical in the three bovine DNA preparations, whereas it is significantly different in each of the other nucleic acids analysed. As the ratios of the other bases are very similar in cow, sheep and rat DNA, the content of methylcytosine appears to be one of the most characteristic differences between nucleic acids from different species.

When more distantly related species are compared the divergences become greater. Herring DNA differs significantly from bovine DNA with respect to guanine as well as methylcytosine, and in locust and sea-urchin nucleic acids the differences are greater again.

Remarkable, however, is the constant pattern of composition in all the animal DNA so far analysed (the 'molar ratios' of adenine and thymine being greater, and those of guanine and cytosine less than unity), in contrast with the widely divergent compositions of microbial DNA's (Visscher, Zamenhof & Chargaff, 1949; Smith & Wyatt, 1951). The only DNA of a higher plant so far analysed, that of wheat germ, has a pattern of composition similar to that of the animal nucleic acids, but is marked by an exceptionally high content of methylcytosine.

These characteristic ratios of the purines and pyrimidines in nucleic acids from different sources are not consistent with any simple theory of nucleic acid composition, but indicate a highly complex structure. In view of the localization of DNA in the chromosomes and the indications that DNA composition is more similar among related than unrelated species, it is tempting to speculate whether DNA composition may bear some direct relation to genetic structure.

There is very little evidence whether nucleic acids, as prepared, are homogeneous, or whether they contain mixtures of different molecular types whose total composition is constant from a given source. If the molecules in a given nucleic acid preparation are all of one type, the proportion of methylcytosine (e.g., 1 nucleotide in 75 in ox, 1 in 53 in herring, 1 in 500 in locusts) sets a minimum repeating unit of structure which, although less than the particle weight estimated for thymus nucleic acid by physical methods ($0{\cdot}5$–2×10^6, e.g., Cecil & Ogston, 1948), implies a remarkably elaborate molecule. Though this may be the case, it is equally possible that the preparation contains a number of particle types which, although differing in composition, are similar in size and net charge and so appear homogeneous when examined by electrophoresis or centrifugation.

SUMMARY

1. A micromethod has been devised for the estimation of the purine and pyrimidine bases in deoxypentose nucleic acids. Both purines and pyrimidines are released by a single hydrolysis with formic or perchloric acid and separated on a single paper chromatogram.

2. Extinction coefficients of the common purines and pyrimidines have been determined at their absorption maxima in the ultraviolet.

3. The composition of deoxypentose nucleic acid is unaltered by treatment with N-sodium hydroxide (Schmidt & Thannhauser, 1945) and this fractionation can satisfactorily be used for obtaining specimens free from ribonucleic acid for analysis.

4. Simple polynucleotides calculated from the estimated purines and pyrimidines account for about 90 % of the total phosphorus and up to 94 % of the total nitrogen of the various deoxypentose nucleic acid preparations analysed.

5. In analyses of deoxypentose nucleic acids from eight animal sources and from one plant the ratios of the purines and pyrimidines vary with the species source of the nucleic acid, but not with different tissues from one species. The content of 5-methylcytosine shows particularly characteristic variations.

I am indebted to the Anti-Locust Research Centre for supplies of locusts, and to Lord Rothschild for samples of bull semen and *Echinus* semen. I wish also to thank Dr R. Markham and Dr J. D. Smith for helpful discussion of the work. The work was carried out while I was on transfer of research from the Forest Insect Investigations Unit, Division of Entomology, Department of Agriculture, Canada, to the Agricultural Research Council Plant Virus Research Unit, Molteno Institute, Cambridge. This paper is Contribution No. 2702, Division of Entomology, Science Service, Department of Agriculture, Ottawa, Canada.

REFERENCES

Allen, R. J. L. (1940). *Biochem. J.* **34**, 858.
Avery, O. T., MacLeod, C. M. & McCarty, M. (1944). *J. exp. Med.* **79**, 137.
Boivin, A. (1947). *Cold. Spr. Harb. Sym. quant. Biol.* **12**, 7.
Cecil, R. & Ogston, A. G. (1948). *J. chem. Soc.* p. 1382.
Chargaff, E., Vischer, E., Doniger, R., Green, C. & Misani, F. (1949). *J. biol. Chem.* **177**, 405.
Chargaff, E., Zamenhof, S. & Green, C. (1950). *Nature, Lond.*, **165**, 756.
Daly, M. M., Allfrey, V. G. & Mirsky, A. E. (1950). *J. gen. Physiol.* **33**, 497.
Hammarsten, E. (1924). *Biochem. Z.* **144**, 383.
Hotchkiss, R. D. (1948). *J. biol. Chem.* **175**, 315.
Levene, P. A. & Bass, L. W. (1931). *Nucleic Acids.* New York: Chemical Catalog Co.
Levene, P. A. & Mandel, J. A. (1908). *Biochem. Z.* **10**, 215.
Markham, R. & Smith, J. D. (1949). *Biochem. J.* **45**, 294.
Marshak, A. & Vogel, H. J. (1950). *Fed. Proc.* **9**, 85.
Mirsky, A. E. & Pollister, A. W. (1946). *J. gen. Physiol.* **30**, 117.
Morton, R. A. (1949). *Photoelectric Spectrometry Group Bull.* **1**, 15.
Overend, W. G. & Webb, M. (1950). *Research*, **3**, 238.
Schmidt, G. & Thannhauser, S. J. (1945). *J. biol. Chem.* **161**, 83.
Smith, J. D. & Markham, R. (1950). *Biochem. J.* **46**, 509.
Smith, J. D. & Wyatt, G. R. (1951). Unpublished observation.
Steudel, H. (1906). *Hoppe-Seyl. Z.* **49**, 406.
Tisdall, F. F. (1922). *J. biol. Chem.* **50**, 329.
Vischer, E. & Chargaff, E. (1948*a*). *J. biol. Chem.* **176**, 703.
Vischer, E. & Chargaff, E. (1948*b*). *J. biol. Chem.* **176**, 715.
Vischer, E., Zamenhof, S. & Chargaff, E. (1949). *J. biol. Chem.* **177**, 429.
Wyatt, G. R. (1950). *Biochem. J.* **47**, vii.
Wyatt, G. R. (1951). *Biochem. J.* **48**, 581.

OCCURRENCE OF 5-METHYL-CYTOSINE IN NUCLEIC ACIDS

By G. R. WYATT

(*Reprinted from Nature, Vol.* 166, *p.* 237, *August* 5, 1950)

Occurrence of 5-Methyl-Cytosine in Nucleic Acids

THE presence in a nucleic acid of the pyrimidine 5-methyl-cytosine was first reported in 1925 by Johnson and Coghill[1], who claimed its discovery among the hydrolysis products of tuberculinic acid. As their identification, however, was based solely on the optical properties of the crystalline picrate, the correctness of this report has been subject to speculation; yet until the recent application of paper chromatography to nucleic acid analysis, there has been no published attempt to confirm their finding. Recently, using a chromatographic method, Vischer, Zamenhof and Chargaff[2] have estimated the purines and pyrimidines in deoxypentose nucleic acid from avian tubercle bacilli, and could find no trace of methyl-cytosine. Hojchkiss[3], however, has noted in hydrolysed thymus nucleic acid a small amount of a substance the chromatographic behaviour and ultra-violet absorption spectrum of which are compatible with its being 5-methyl-cytosine.

Taking advantage of the sensitive photographic technique of Markham and Smith[4] for the detection of ultra-violet-absorbing substances on paper chromatograms, I have now been able to demonstrate and estimate 5-methyl-cytosine in several animal and plant deoxyribonucleic acids, but not in certain viral and bacterial nucleic acids, including that of tubercle bacilli.

The bases liberated by hydrolysis of nucleic acids with formic acid (175° C. for 30 min.) are separated on one-dimensional paper chromatograms using as the solvent an aqueous solution containing 65 per cent *iso*-propanol and 2·0 *N* hydrochloric acid. In this system, guanine, adenine, cytosine and thymine are separated in that order, and it was noted first in chromatograms of nucleic acid from herring sperm that a faint additional spot is visible just beyond cytosine, which was also present when hydrochloric acid was used for hydrolysis. A sample of this substance, isolated by eluting from large sheets of filter paper, was found to give no colour with ninhydrin and to withstand prolonged acid hydrolysis. It moved just ahead of cytosine in all chromatographic solvents tried, whether acid, neutral or ammoniacal.

This suggested that the new substance might be methyl-cytosine. Treatment with nitrous acid gave a product identical with thymine in its chromatographic behaviour and ultra-violet spectrum. Of the three substances which would yield thymine on

deamination, the unknown was clearly different from 5-methyl-*iso*-cytosine (2-amino-4-hydroxy-5-methyl-pyrimidine, kindly provided by Prof. A. R. Todd), and was unlikely to be 2,4-diamino-5-methyl-pyrimidine in view of the effect an additional amino-group would have on chromatographic movement. It therefore appeared to be 2-hydroxy-4-amino-5-methyl-pyrimidine or 5-methyl-cytosine.

Recently, through the kindness of Mr. A. Pircio, of Fordham University, New York, I have obtained a sample of synthetic 5-methyl-cytosine, and have confirmed its identity with the natural substance by chromatography and by ultra-violet spectrometry. Absorption maxima occur at the following wavelengths: at pH 7, 274 μ; in 0·1 *N* hydrochloric acid, 212 and 283 μ; in 0·1 *N* sodium hydroxide, 287 μ.

Other available nucleic acids were examined for this new pyrimidine. By eluting from chromatograms run in *iso*-propanol-water – hydrochloric acid, concentrating the eluate, and chromatographing in butanol – water – ammonia, amounts of it as small as 1 per cent of the cytosine present (less than 0·1 per cent of the nucleic acid) could be detected and estimated. The following is a list of nucleic acids examined, all of which, except those noted, are of the deoxypentose type.

Source of nucleic acid	Methyl-cytosine content as molar % of cytosine
Calf thymus	6
Beef spleen	6
Ram sperm	5
Herring sperm	8
Locusta migratoria (whole)	1
Wheat germ	33
Tubercle bacilli, human and bovine	0
Escherichia coli	0
E. coli bacteriophages T_2 and T_4	0
Virus of gypsy moth polyhedral disease	0
Yeast (ribose type)	0
Turnip yellow mosaic virus (ribose type)	0

So far, 5-methyl-cytosine has been found only in deoxypentose nucleic acids from animals and higher plants, which contain relatively small amounts characteristic of the source. I am unable to confirm the claim of Rosedale[5] that in the nucleic acids of locusts and certain other insects thymine is quantitatively replaced by 5-methyl-cytosine.

The question arises whether 5-methyl-cytosine is a constituent of the deoxyribonucleic acid macromolecule or of a smaller molecule, perhaps another nucleic acid, mixed in the preparations. Herring sperm nucleic acid fractionated by ultra-centrifugation showed no difference in composition of sediment and supernatant. Nor is its composition altered by precipitation after depolymerization with 1 *N* sodium hydroxide. On the assumption that it consists of a single molecular species, the locust nucleic acid analysed would contain approximately one nucleotide of methyl-cytosine in every 400, corresponding to a molecular weight of about 150,000. In view of the evident complexity and specificity of nucleic acids, there seems to be no reason why this should not be so; yet a further attempt at fractionation would be worth while.

The thymus nucleic acid used was the gift of Dr. K. Bailey; ram sperm nucleic acid was prepared in collaboration with Dr. T. Mann; dried tubercle bacilli were provided by the Wellcome Research Laboratories, Ltd. I am grateful to Dr. K. M. Smith for providing facilities for this work, to Dr. Roy Markham for help and advice, and to Dr. J. D. Smith for preparation and analysis of nucleic acids from *E. coli*, bacteriophages and turnip yellow mosaic virus. The work was carried out while I was on transfer of research from the Division of Entomology, Department of Agriculture, Ottawa, Canada.

G. R. Wyatt

Agricultural Research Council,
Plant Virus Research Unit,
Molteno Institute,
Cambridge.

April 3.

[1] Johnson, T. B., and Coghill, R. D., *J. Amer. Chem. Soc.*, **47**, 2838 (1925).
[2] Vischer, E., Zamenhof, S., and Chargaff, E., *J. Biol. Chem.*, **177**, 429 (1949).
[3] Hotchkiss, R. D., *J. Biol. Chem.*, **175**, 315 (1948).
[4] Markham, R., and Smith, J. D., *Biochem. J.*, **45**, 294 (1949).
[5] Rosedale, J. L., *J. Ent. Soc. S. Africa*, **11**, 34 (1948).

Printed in Great Britain by Fisher, Knight & Co., Ltd., St. Albans.