THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 236, No. 3, March 1961
Printed in U.S.A.

Enzymatic Synthesis of Deoxyribonucleic Acid

VIII. FREQUENCIES OF NEAREST NEIGHBOR BASE SEQUENCES IN DEOXYRIBONUCLEIC ACID

JOHN JOSSE,* A. D. KAISER, AND ARTHUR KORNBERG

From the Department of Biochemistry, Stanford University School of Medicine, Palo Alto, California

(Received for publication, October 4, 1960)

Determination of deoxyribonucleotide sequence in a deoxyribonucleic acid molecule is important from both the chemical and genetic points of view. It is also essential for answering the question of whether DNA synthesized *in vitro* by polymerase (1, 2) is a faithful copy of the nucleotide[1] sequence of the primer DNA. Although enzymatically synthesized DNA has the same over-all nucleotide composition as the particular primer DNA (3), it could not be inferred that this synthesis is a replication of the nucleotide sequences of the primer.

Because of the limitations of present methods, complete sequence studies have never been made. Sinsheimer (4) has been able to identify 16 of the 25 possible dinucleotide pairs resulting from enzymatic digestion of calf thymus DNA. From recent analyses of various DNA's by Shapiro and Chargaff (5) and by Burton and Peterson (6), there is now information on the frequency of occurrence of a pyrimidine nucleotide flanked on both sides by purine nucleotides and on the frequencies of short runs of pyrimidine nucleotides in the polynucleotide chains. These workers were able to deduce that there is a nonrandom arrangement of pyrimidine nucleotides in all of the DNA's examined.

In the studies to be reported here, we have derived the frequencies of the 16 possible nearest neighbor pairs in a variety of DNA's by the technique of enzymatic incorporation of 5′-P^{32}-labeled nucleotides into DNA and then degradation of the DNA into 3′-nucleotides. Briefly, we have found that: (*a*) each DNA directs the synthesis of a product which has a unique and nonrandom pattern of the 16 nearest neighbor frequencies; (*b*) the DNA synthesized has the same nearest neighbor frequencies whether the primer is native DNA or enzymatically prepared DNA containing only traces of the original native DNA; and (*c*) the pattern of nearest neighbor frequencies in every case involves both base-pairing of adenine to thymine and of guanine to cytosine between sister strands of DNA, and opposite "polarity" of the two strands as proposed in the Watson and Crick model (7).

EXPERIMENTAL PROCEDURE

Materials

Substrates and Enzymes—Unlabeled deoxynucleoside triphosphates were synthesized by the method of Smith and Khorana (8). All of the labeled substrates contained P^{32} in the phosphate esterified to the sugar; they were prepared as described previously (1). The DNA-synthesizing enzyme was prepared from the polymerase, Fraction VII, described elsewhere (1); this enzyme was refractionated with diethylaminoethyl cellulose, yielding a preparation with a specific activity of 500 units per mg of protein. Micrococcal DNase was prepared according to Cunningham *et al.* (9); the final fraction had a specific activity of 7500 units per mg of protein.[2] Calf spleen phosphodiesterase was isolated by Hilmoe's procedure (10); the purified preparation had a specific activity of 33 units per mg of protein. It is important that the latter two enzymes be entirely free of phosphomonoesterase activity. A crude semen phosphomonoesterase was prepared from human semen by ammonium sulfate precipitation (11); the preparation had a specific activity of 1200 units per mg of protein.[3] A phosphodiesterase from *Escherichia coli*, kindly supplied by Dr. I. R. Lehman, was the concentrated diethylaminoethyl cellulose fraction and had a specific activity of 670 units per mg of protein (13).

DNA Preparations—Samples from *E. coli*, *Aerobacter aerogenes*, *Mycobacterium phlei*, and *Hemophilus influenzae* were prepared by a method previously described (3). *Micrococcus lysodeikticus* DNA was purified by the same procedure from cell extracts made by lysozyme treatment. *Bacillus subtilis* DNA was generously provided by Dr. E. W. Nester; it had been extracted from the cells by lysozyme treatment, deproteinized by the Sevag method (14), and then treated with pancreatic RNase (Worthington Biochemical Corporation). DNA's from the bacteriophages T2r+, T4r+, and T6r+ were extracted from phages grown and purified by the method of Herriott and Bar

* Fellow of the National Foundation.

[1] The abbreviations used are: nucleotide, deoxyribonucleotide; $dATP^{32}$, $dTTP^{32}$, $dCTP^{32}$, $dGTP^{32}$, triphosphates containing P^{32} at the phosphate esterified to the sugar; d-AT copolymer, copolymer of deoxyadenylate and deoxythymidylate; d-GC polymer, polymer of polydeoxyguanylate and polydeoxycytidylate.

[2] The enzyme was assayed by incubating 10 μmoles of Tris buffer (pH 8.6), 1 μmole of $CaCl_2$, 1 μmole of native calf thymus DNA (expressed as nucleotide P equivalents), and enzyme in a volume of 0.3 ml. After 15 minutes at 37°, the reaction was stopped by the addition of 1.0 ml of cold water and 1.0 ml of cold 1 N perchloric acid. After 5 minutes at 0°, the precipitate was removed by centrifugation, and the optical density of the supernatant liquid at 260 mμ was determined. A unit of enzyme was defined as the amount causing the production of 10 mμmoles of acid-soluble DNA polynucleotides in 15 minutes (assuming a molar extinction coefficient of 10,000).

[3] A unit of enzyme was defined as the amount causing liberation of 1 μmole of inorganic phosphate in 1 hour in an assay mixture consisting of 0.1 mmole of sodium acetate (pH 5.0), 10 μmoles of $MgCl_2$, 3 μmoles of 5′-AMP, and enzyme in a final volume of 1.2 ml. After 15 minutes at 37° the reaction was stopped by the addition of 0.14 ml of 3 M trichloroacetic acid. An 0.8-ml aliquot was removed for inorganic phosphate analysis by the Fiske and SubbaRow method (12).

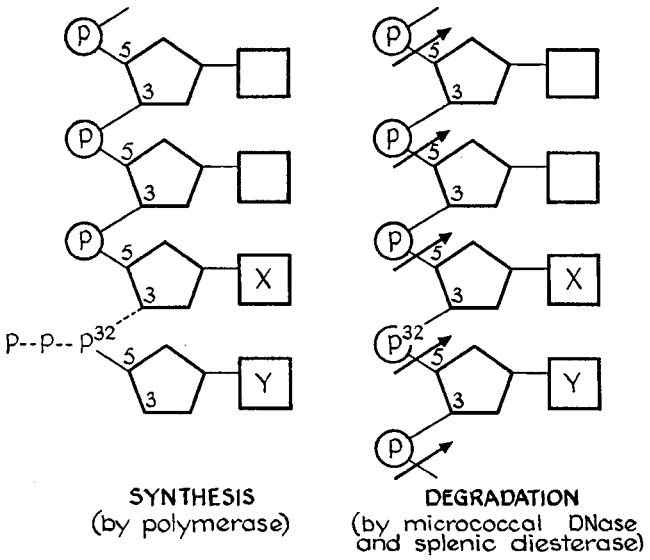

FIG. 1. Synthesis of a P^{32}-labeled DNA chain and its subsequent enzymatic degradation to 3′-deoxyribonucleotides. The arrows indicate the linkages cleaved by micrococcal DNase and calf spleen phosphodiesterase, yielding a digest composed exclusively of 3′-deoxyribonucleotides.

low (15). After osmotic shock, residual intact bacteriophages and ghosts were removed by centrifugation; residual protein in the supernatant fluid was removed by the Sevag method (14), after which the DNA was precipitated by the addition of 2 volumes of cold 95% ethanol. DNA's from the temperate bacteriophages λ and λdg were prepared by phenol treatment of phages purified in a CsCl density gradient (16).

Calf thymus DNA was isolated by the method of Kay, Simmons, and Dounce (17). Heated calf thymus DNA was prepared by heating at 100° for 30 minutes a neutral solution containing 0.5 mg of DNA per ml of water; under these conditions a 39% hyperchromic effect was demonstrated. Enzymatically synthesized "calf thymus" DNA was prepared by incubating native calf thymus DNA, the four unlabeled deoxynucleoside triphosphates, and polymerase of *E. coli* as previously described (2); the reaction was allowed to proceed for 6 hours and then the solution was dialyzed for 48 hours at 5° against 5000 volumes of 0.2 M NaCl. The nondialyzable product now contained 20 times the optical density at 260 mμ of the primer DNA. Thus, 95% of the DNA in the product was derived from the substrates by enzymatic synthesis.

d-AT copolymer was prepared as described by Schachman *et al.* (18). P^{32}-labeled d-AT copolymer was prepared with $dATP^{32}$. The d-GC polymer was enzymatically synthesized by incubation of dGTP and dCTP with polymerase in the absence of added DNA primer; after a lag period of 1 to 2 hours, a viscous product containing deoxyguanylate and deoxycytidylate was formed. Further details concerning the synthesis and characterization of the d-GC polymer will be reported at a future date.

Methods

I. Basic Plan

Synthesis of P^{32}-labeled DNA's—Each DNA used as primer was reacted with labeled substrates according to the following pattern:

Reaction 1. $dATP^{32}$, dTTP, dGTP, dCTP
Reaction 2. dATP, $dTTP^{32}$, dGTP, dCTP
Reaction 3. dATP, dTTP, $dGTP^{32}$, dCTP
Reaction 4. dATP, dTTP, dGTP, $dCTP^{32}$.

As shown in Fig. 1, the P^{32}, which is attached to carbon 5 of the deoxyribose of the nucleoside triphosphate (Y), becomes esterified with carbon 3 of the deoxyribose of the nucleotide (X) at the growing end of the chain. Approximately 1 mμmole of P^{32}-labeled diester bonds, representing of the order of 10^{15} such linkages, was made in each reaction mixture. (Although this mechanism for chain growth appears the most probable, alternative mechanisms will not alter the P^{32} distribution in the synthesis of long chains.) The extent of DNA synthesis generally represented a 20% increment over the amount of primer added, but, as will be apparent from the calculations, it was not essential that the reactions in each of the four mixtures proceed to the same extent. It is assumed that the pattern of nucleotide arrangements will not be different whether the net increase in DNA varies, for example, from 10 to 30%.

Degradation of P^{32}-labeled DNA's—The DNA isolated and washed free of unaltered substrates was hydrolyzed by the consecutive action of micrococcal DNase and calf spleen phosphodiesterase. These enzymes cleave the bonds between phosphate and carbon 5 of deoxyribose, leaving as products 3′-mononucleotides. Thus, the P^{32} now labels the deoxyribose carbon 3 of that nucleotide in the chain with which the labeled triphosphate substrate reacted (Fig. 1). By determining the P^{32} content of each of the four 3′-mononucleotides isolated from each reaction digest, one can calculate the frequency with which any nucleotide is linked to another in the newly synthesized chains.

II. Detailed Procedure

A. Enzymatic Synthesis of P^{32}-labeled DNA—Each incubation mixture (0.3 ml) contained 20 μmoles of glycine buffer (pH 9.2), 2 μmoles of $MgCl_2$, 0.3 μmole of 2-mercaptoethanol, 20 mμmoles of a DNA primer (expressed as nucleotide P equivalents), 5 mμmoles each of dATP, dTTP, dGTP, and dCTP, only one of which was labeled with P^{32} (specific activity of 0.5 to 1.0 × 10^8 c.p.m. per μmole), and polymerase. The amount of enzyme used was sufficient to incorporate approximately 1 mμmole of the labeled nucleotide into DNA in 30 minutes. The amount of enzyme necessary varied with the different DNA primers used and was determined by preliminary assays with each type of primer; 0.5 to 5.0 μg of enzyme protein were required. For each primer, four such incubation mixtures were prepared, each containing the same amount of enzyme but a different labeled triphosphate, as described under the basic plan; all the reactions ran for 30 minutes at 37°.

B. Isolation of Enzymatically Synthesized DNA—At the end of the incubation, each mixture was chilled; the isolation steps were carried out at 2°. Calf thymus DNA (0.2 ml of a solution containing 5 μmoles of nucleotide P per ml) was added as "carrier" followed at once by addition of 0.5 ml of 1 N perchloric acid. After 5 minutes, 2.5 ml of water were added, and the precipitate was dispersed. The precipitate was collected by centrifugation, dissolved in 0.3 ml of 0.2 N NaOH, and then reprecipitated with 0.4 ml of 1 N perchloric acid. Water (3.0 ml) was added and the suspension stirred and centrifuged. The precipitate was treated once more in the same way, then collected and dispersed in 0.4 ml of water, and dissolved with the aid of 0.1 N NaOH. The pH was now adjusted to 8.6 and the volume to 0.5 ml

TABLE I

Radioactivity measurements in experiment with M. phlei DNA as primer

Isolated 3′-deoxyribonucleotide	Labeled triphosphate											
	Reaction No. 1. dATP^{32}			Reaction No. 2. dTTP^{32}			Reaction No. 3. dGTP^{32}			Reaction No. 4. dCTP^{32}		
	Sequence	C.p.m.	Fraction	Sequence	C.p.m.	Fraction	Sequence	C.p.m.	Fraction	Sequence	C.p.m.	Fraction
Tp	TpA	873	0.075	TpT	1,665	0.157	TpG	3,490	0.187	TpC	4,130	0.182
Ap	ApA	1,710	0.146	ApT	2,065	0.194	ApG	2,500	0.134	ApC	4,300	0.189
Cp	CpA	4,430	0.378	CpT	2,980	0.279	CpG	7,730	0.414	CpC	6,070	0.268
Gp	GpA	4,690	0.401	GpT	3,945	0.370	GpG	4,960	0.265	GpC	8,200	0.361
Sum		11,703	1.000		10,655	1.000		18,680	1.000		22,700	1.000

Less than 0.1% of the unaltered triphosphates remained in the final precipitate, as judged from control experiments that lacked enzyme or primer DNA.

C. Enzymatic Digestion of DNA to 3′-Nucleotides—Stage 1. To each DNA solution were added 2 μmoles of Tris buffer (pH 8.6), 1 μmole of $CaCl_2$, and 180 units of micrococcal DNase (total volume, 0.54 ml). After incubation at 37° for 2 hours, all the DNA was in the form of acid-soluble fragments.

Stage 2. The pH of each digest was reduced to 7.0 with 0.1 N HCl, 0.13 unit of calf spleen phosphodiesterase was added, and the mixture was incubated at 37°. Another 0.13 unit of the phosphodiesterase was added after 1 hour and again after a second hour. Total incubation time was 3 hours. At this point, aliquots were removed from each of the reaction mixtures and assayed for completeness of digestion, as judged by the fraction of P^{32} which had become susceptible to the action of phosphomonoesterase.[4] In all 72 digests, over 93% of the radioactivity was phosphomonoesterase-sensitive; in 68 of the digests, it was greater than 95%. The digest was then dried under an air stream at room temperature and dissolved in 0.07 ml of water.

D. Separation of 3′-Nucleotides—Each preparation was subjected to paper electrophoresis according to the technique of Markham and Smith (19) (Whatman No. 3MM paper, 0.05 M ammonium formate buffer at pH 3.5, 1200 volts, 25–35°, 2 to 2½ hours). The bands were identified under an ultraviolet lamp, cut out, and eluted overnight at room temperature into 3 ml of 0.01 N HCl. The identity of the 3′-nucleotide bands was occasionally verified spectrophotometrically. Strips of paper between the sharply resolved bands and in the forward area where inorganic phosphate travels were also cut out and eluted.

E. Determination of P^{32} Content of 3′-Nucleotides—The eluates (1.0-ml aliquots) were assayed for radioactivity in a Nuclear-Chicago model D-47 gas flow counter with Micromil window. In all 72 digests, over 93% of the radioactivity applied to the paper was recovered in the nucleotide bands; in 65 of the digests, recovery was over 95%. Little or no radioactivity (<2%) was found in the strips between the mononucleotide bands or in the inorganic phosphate band, confirming that digestion of the DNA had been complete, and there had been no hydrolysis of the mononucleotides.

[4] The assay mixture (0.2 ml) consisted of a 0.03-ml aliquot added to 50 μmoles of sodium acetate buffer (pH 5.0), 3 μmoles of $MgCl_2$, and 70 units of human semen phosphomonoesterase. After 20 minutes of incubation at 37°, the assay mixture was chilled and treated exactly as described previously in Stage III of the assay for deoxynucleotide kinase (1). Radioactivity still adsorbed to Norit indicated its presence in a form resistant to the action of phosphomonoesterase.

RESULTS

Calculation of Relative Frequencies of Nearest Neighbor Nucleotide Sequences—Because the specific radioactivities of the labeled substrates are known, it should be possible by measuring the radioactivity in each of the four 3′-nucleotides from a reaction digest to determine directly the amounts of these nucleotides. However, the variability in the progress of the enzymatic synthesis in the four reaction mixtures and in the isolation of the synthesized DNA reduced the precision of the analysis and led to the use of the method described below. The experiment with DNA from *M. phlei* as primer will be used as an example (Table I).

The P^{32} content of each of the four 3′-mononucleotides in a given reaction digest is converted to the decimal fraction of the total. The fractional values in a given reaction are not dependent upon the particular synthetic conditions in that reaction. In fact, identical fractional values should be (and indeed have been) obtained with varying amounts of substrates, enzyme, and incubation times. However, the relative amounts of each of the four bases incorporated into the enzymatically synthesized DNA are known to vary, depending upon the base composition of the particular primer DNA (3). It is therefore necessary to multiply the fractional values of each reaction by a base-incorporation factor, which expresses the relative frequency with which the particular base (originally labeled as a 5′-nucleotide) is incorporated into the synthesized DNA. The four base-incorporation factors, which we have denoted a, t, g, and c, for the labeled nucleoside triphosphates of adenine, thymine, guanine, and cytosine, respectively, may be derived by either of two methods.

One method is to use the values provided by chemical determination of the base composition of the primer DNA, inasmuch as it has already been established that the base composition of enzymatically synthesized DNA is identical to that of the particular primer (3). In the case of *M. phlei* primer DNA, the molar proportions of adenine, thymine, guanine, and cytosine are 0.162, 0.165, 0.338, and 0.335, respectively (3), and therefore the base-incorporation factors are $a = 0.162$, $t = 0.165$, $g = 0.338$, and $c = 0.335$.

The preferred method, which has been routinely used, requires no independent knowledge of the base composition of the particular primer DNA. Instead, the base-incorporation factors

are derived from the radioactivity measurements in the experiment. In these experiments, the digestions of the DNA and the subsequent recoveries of 3′-nucleotides were complete, and as a consequence the amount of a given base recovered from the synthesized DNA as a 3′-nucleotide is equal to the amount of that base incorporated into the DNA during synthesis as a 5′-nucleotide. Four equations involving a, t, g, and c and the 16 fractional values of Table I can be written expressing this equivalence. For example, the total amount of adenine incorporated as a 5′-nucleotide (TpA + ApA + CpA + GpA) must be equal to the total amount of adenine recovered as a 3′-nucleotide (ApA + ApT + ApG + ApC). For *M. phlei* DNA (Table I) the equation expressing this equivalence is:

$$0.075\,a + 0.146\,a + 0.378\,a + 0.401\,a = 0.146\,a + 0.194\,t + 0.134\,g + 0.189\,c$$

$$a = 0.146\,a + 0.194\,t + 0.134\,g + 0.189\,c$$

Similarly for thymine, guanine, and cytosine nucleotides:

$$t = 0.075\,a + 0.157\,t + 0.187\,g + 0.182\,c$$
$$g = 0.401\,a + 0.370\,t + 0.265\,g + 0.361\,c$$
$$c = 0.378\,a + 0.279\,t + 0.414\,g + 0.268\,c$$

The solution to these 4 equations is:

$$a = 0.489\,c; \quad t = 0.483\,c; \quad g = 1.000\,c.$$

It is apparent that the data show incorporation of deoxyadenylate equal to deoxythymidylate and incorporation of deoxyguanylate equal to deoxycytidylate. The ratio $(a + t)/(g + c) = 0.48$ is the same value as obtained by chemical analysis of the *M. phlei* DNA primer (3).

For the convenience of expressing the frequencies of each of the nearest neighbor sequences as a decimal proportion of one, we may set:

$$a + t + g + c = 1.000.$$

Then substituting the values derived above we obtain:

$$a = 0.164, \quad t = 0.162, \quad g = 0.337, \quad c = 0.337.$$

The base-incorporation factors derived algebraically from the radioactivity measurements are in the case of *M. phlei* DNA seen to be identical to those available from chemical determination of base composition. On the one hand, this agreement validates the assumptions entailed in the algebraic derivation of the base-incorporation factors. On the other hand, the results indicate that the derived base-incorporation factors are accurate expressions of the nucleotide composition of both primer and product DNA's.

The decimal fractions given in Table I may now be properly weighted for base frequencies, with each multiplied by its base-incorporation factor, and the nearest neighbor frequencies obtained are shown in Table II. As can be seen in Table II, the sums of the nearest neighbor frequencies for each nucleotide are the respective base-incorporation factors.

Another example of the calculation of nearest neighbor frequencies from the original data is the experiment in which DNA from *H. influenzae*, a primer with relatively high adenine and thymine content, was used (Tables III and IV). The derived base-incorporation factors for this DNA, revealed in the nucleotide sums of Table IV, are, as in the case of the *M. phlei* DNA, very similar to those obtained by independent chemical analysis of the primer DNA.

Base Composition of Enzymatically Synthesized DNA—The nucleotide composition of enzymatically synthesized DNA, indicated in the nucleotide sums of the nearest neighbor frequencies, agreed with the primer compositions in *M. phlei* and *H. influenzae*, the examples mentioned earlier. Data obtained in the same way from nearest neighbor frequency determinations on a variety of DNA primers are summarized in Table V. The results indicate that in each of these experiments faithful replication of the over-all composition of the primer DNA has been achieved. Furthermore, the results independently confirm earlier evidence from chemical analyses that the base composition of enzymatically synthesized DNA is identical to that of the primer DNA (3).

Base-pairing and Opposite Polarity of Strands in Enzymatically Synthesized DNA—In the Watson and Crick model of DNA (7), two polynucleotide strands are held together in a helix by hydrogen bonds between adenine and thymine and between guanine and cytosine. It was proposed that each strand could serve as a template for the formation of a new polynucleotide chain, the alignment of nucleotides in the new chain being mediated by the same hydrogen-bonding forces (Fig. 2). The postulated scheme allows for precise replication, since one parent helix gives rise to two daughter helices identical to one another and to the parent molecule. Replication of base composition of the primer in the enzymatic synthesis of DNA was presented as evidence for this scheme (3). Now analysis of the frequencies of nearest neighbor sequences provides independent support for this mechanism.

The tables show that the amounts of ApA and of TpT sequences are equivalent; and so are the frequencies of CpC and GpG sequences (Tables II and IV). In examining the pairing of the other sequences, it is essential to recognize that the results expected are distinctly different, depending upon whether the strands of the double helix are of similar or opposite polarity. For example, the short segments illustrated in Fig. 3 contrast strands of similar polarity with strands of opposite polarity. The matching sequences predicted by the two models are different.

In the model with opposite polarity six matching sequences are predicted; they are indicated in Tables II and IV by the same roman numeral. In each instance the agreement is good. The four values along the diagonal, which separates the data into two symmetrical halves, are independent and cannot be checked. In other words, every TpA sequence would be matched by a TpA sequence in the complementary strand of opposite polarity; the same constraint would apply to ApT, CpG, and GpC sequences.

In the model with strands of similar polarity the 16 nearest neighbor sequence frequencies would fall into 8 pairs of matching values, indicated in Tables II and IV by the same lower case letter. Excluding the ApA, TpT, CpC, and GpG sequences, which match similarly in both models, it is evident that in only 4 out of 12 instances are the values reasonably close. Statistical analysis of the data confirms good fit to the model of opposite polarity but significant deviation from the model of similar polarity (Tables VI, VII, and VIII). As can be seen, the σ_{opp}, which is the standard deviation of those values predicted to be equal in a model with strands of opposite polarity, is in every experiment close to the σ_{error}, the estimated standard deviation of values due to the over-all analytical error of the method ($\sigma_{opp}/\sigma_{error}$ ratios near 1.0). On the other hand, the σ_{sim}, which is

TABLE II

*Nearest neighbor frequencies of M. phlei DNA**

As described in the text, identical Roman numerals designate those sequence frequencies which should be equivalent in a Watson and Crick DNA model with strands of opposite polarity; identical lower case letters designate sequence frequencies which should be equivalent in a model with strands of similar polarity.

Reaction No.	Labeled triphosphate	Isolated 3'-deoxyribonucleotide			
		Tp	Ap	Cp	Gp
1	dATP^{32}	a TpA 0.012 I	b ApA 0.024	c CpA 0.063 II	d GpA 0.065 III
2	dTTP^{32}	b TpT 0.026 I	a ApT 0.031	d CpT 0.045 IV	c GpT 0.060 V
3	dGTP^{32}	e TpG 0.063 II	f ApG 0.045 IV	g CpG 0.139	h GpG 0.090 VI
4	dCTP^{32}	f TpC 0.061 III	e ApC 0.064 V	h CpC 0.090 VI	g GpC 0.122
	Sums	0.162	0.164	0.337	0.337

* Chemical analysis of the base composition of the primer DNA indicated molar proportions of thymine, adenine, cytosine, and guanine of 0.165, 0.162, 0.335, and 0.338, respectively (3).

TABLE III

Radioactivity measurements in experiment with H. influenzae DNA as primer

Isolated 3'-deoxy-ribonucleotide	Labeled triphosphate											
	Reaction No. 1. dATP^{32}			Reaction No. 2. dTTP^{32}			Reaction No. 3. dGTP^{32}			Reaction No. 4. dCTP^{32}		
	Sequence	C.p.m.	Fraction	Sequence	C.p.m.	Fraction	Sequence	C.p.m.	Fraction	Sequence	C.p.m.	Fraction
Tp	TpA	8,406	0.234	TpT	14,196	0.377	TpG	7,848	0.354	TpC	6,633	0.273
Ap	ApA	13,404	0.373	ApT	11,610	0.309	ApG	5,805	0.262	ApC	6,246	0.257
Cp	CpA	7,800	0.217	CpT	5,985	0.159	CpG	4,308	0.194	CpC	4,713	0.194
Gp	GpA	6,327	0.176	GpT	5,841	0.155	GpG	4,209	0.190	GpC	6,696	0.276
Sum		35,937	1.000		37,632	1.000		22,170	1.000		24,288	1.000

the standard deviation of those values predicted to be equal in a model with strands of similar polarity, is in every experiment considerably higher than the σ_{error} ($\sigma_{sim}/\sigma_{error}$ ratios greater than 4.7).

Nonrandom Nature of Nearest Neighbor Frequencies in Enzymatically Synthesized DNA—Because there are about 10^4 nearest neighbor pairs in a DNA strand and perhaps 10^8 or more different strands even in a bacterial DNA population, it was possible

TABLE IV

*Nearest neighbor frequencies of H. influenzae DNA**

See legend of Table II.

Reaction No.	Labeled triphosphate	Isolated 3′-deoxyribonucleotide			
		Tp	Ap	Cp	Gp
1	dATP^{32}	a TpA 0.073	b ApA 0.116 I	c CpA 0.067 II	d GpA 0.054 III
2	dTTP^{32}	b TpT 0.116 I	a ApT 0.095	d CpT 0.049 IV	c GpT 0.048 V
3	dGTP^{32}	e TpG 0.067 II	f ApG 0.050 IV	g CpG 0.038	h GpG 0.036 VI
4	dCTP^{32}	f TpC 0.052 III	e ApC 0.049 V	h CpC 0.037 VI	g GpC 0.053
	Sums	0.308	0.310	0.191	0.191

* Chemical analysis of the base composition of the primer DNA indicated molar proportions of thymine, adenine, cytosine, and guanine of 0.302, 0.319, 0.196, and 0.182, respectively (20).

TABLE V

Nucleotide composition of enzymatically synthesized DNA

The values in the first 4 columns are taken from the nucleotide sums of the nearest neighbor frequencies of each DNA; in the case of the bacteriophage DNA values, averages of the λ and λ dg sums and of the T2, T4, and T6 sums are given in this table.

Primer DNA	Ap	Tp	Gp	Cp	$\frac{Ap + Gp}{Tp + Cp}$	$\frac{Ap + Tp}{Gp + Cp}$	Chemical analysis of primer $\frac{A + T^*}{G + C}$	Reference
M. lysodeikticus	0.147	0.145	0.354	0.354	1.00	0.41	0.39	21
M. phlei	0.164	0.162	0.337	0.337	1.00	0.48	0.49	3
A. aerogenes	0.220	0.223	0.280	0.277	1.00	0.80	0.82	3
E. coli	0.248	0.254	0.249	0.249	0.99	1.01	0.97	3
Bacteriophages λ and λ dg	0.247	0.252	0.249	0.252	0.99	1.00	1.06	16
B. subtilis	0.278	0.280	0.222	0.220	1.00	1.26	1.29	†
Calf thymus	0.286	0.283	0.214	0.217	1.00	1.32	1.25	3
H. influenzae	0.310	0.308	0.191	0.191	1.00	1.62	1.64	20
Bacteriophages T2, T4 and T6	0.319	0.318	0.184	0.179	1.01	1.76	1.84‡	22, 23
d-AT copolymer	0.500	0.500	<0.002	<0.002	1.00	>250	>40	3

* A, T, G, and C refer to the bases adenine, thymine, guanine, and cytosine, respectively.

† Analysis of *B. subtilis* DNA primer was performed according to the method of Wyatt and Cohen (22).

‡ The value for the T-even bacteriophage DNA's represents an average of all the data given in the references.

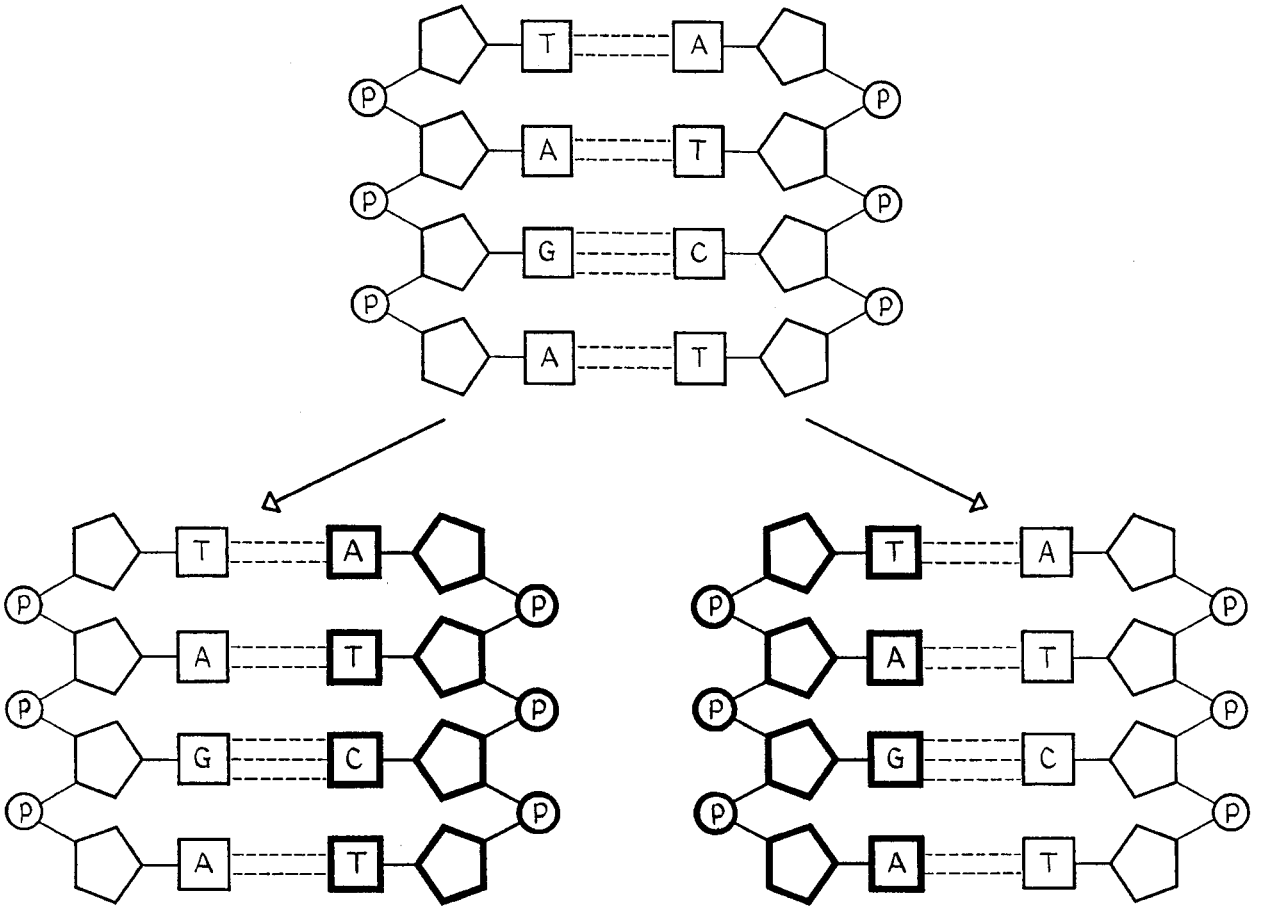

FIG. 2. Proposed scheme of replication of a Watson and Crick DNA model. Bold-lined polynucleotide chains of the two daughter molecules represent newly synthesized strands.

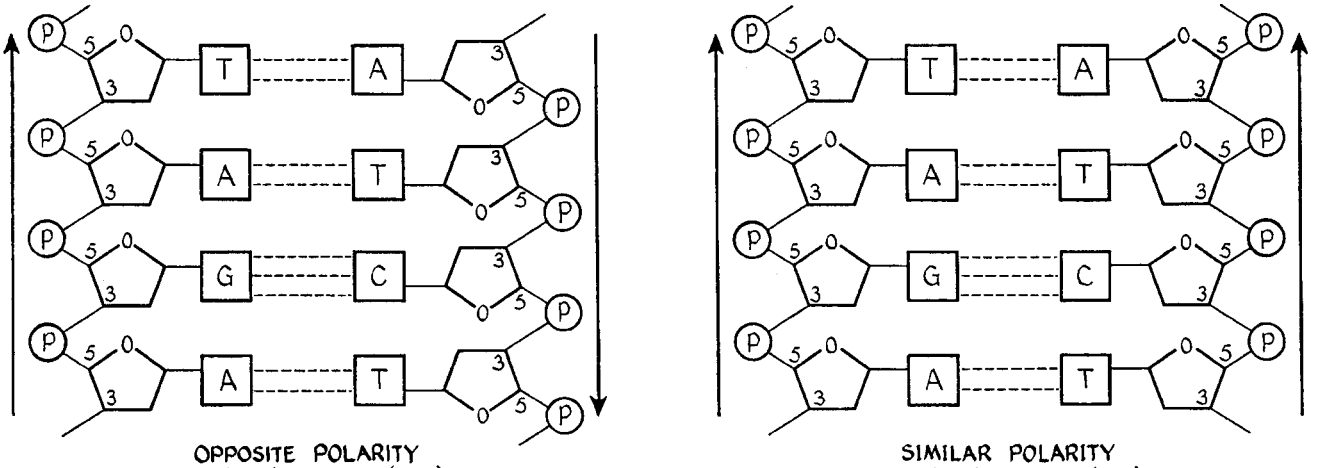

FIG. 3. Contrast of a Watson and Crick DNA model with strands of opposite polarity with a model with strands of similar polarity. The predicted matching nearest neighbor sequence frequencies are different. Values in parentheses are sequence frequencies from the experiment with *M. phlei* DNA. (The strands represented here are the newly synthesized strands of Fig. 2; for ease of comparison, they are aligned as if they were complementary strands of the same double helix.)

TABLE VI

Nearest neighbor frequencies of calf thymus DNA

Nearest neighbor sequence	1	2	3	4
	Native calf thymus DNA	Enzymatically synthesized "calf thymus" DNA*		Heated calf thymus DNA
		Untreated	Treated with phosphodiesterase†	
ApA, TpT	0.089, 0.087	0.093, 0.091	0.088, 0.083	0.087, 0.082
CpA, TpG	0.080, 0.076	0.073, 0.073	0.078, 0.076	0.079, 0.077
GpA, TpC	0.064, 0.067	0.064, 0.059	0.063, 0.064	0.065, 0.064
CpT, ApG	0.067, 0.072	0.067, 0.070	0.068, 0.074	0.070, 0.072
GpT, ApC	0.056, 0.052	0.050, 0.052	0.056, 0.051	0.054, 0.055
GpG, CpC	0.050, 0.054	0.047, 0.054	0.057, 0.055	0.054, 0.054
TpA	0.053	0.076	0.059	0.053
ApT	0.073	0.091	0.075	0.070
CpG	0.016	0.010	0.011	0.017
GpC	0.044	0.039	0.042	0.047
Deviation ratios‡				
$\sigma_{opp}/\sigma_{error}$	1.9	0.8	1.6	0.7
$\sigma_{sim}/\sigma_{error}$	9.3	9.1	9.6	8.9
$\sigma_{rand}/\sigma_{error}$	9.1	7.9	9.4	8.8

* Only 5% of this DNA is the native calf thymus primer; its synthesis is described in "Materials."

† After synthesis of the P^{32}-labeled DNA, the four reaction mixtures were made 0.2 M in NaCl and then heated at 70° for 5 minutes to destroy enzymatic activity. This was followed by dialysis for 48 hours at 5° against 5000 volumes of 0.2 M NaCl and then further dialysis for 4 hours against 1000 volumes of 0.06 M glycine (pH 9.7) at 5°. To each dialysate (2.5 ml) were added 14 μmoles of $MgCl_2$ and 150 units of *E. coli* phosphodiesterase. After 60 minutes at 37°, each reaction mixture was chilled and then treated as in Step B and succeeding steps of the detailed procedure. The nearest neighbor frequencies shown in column 3 were then obtained.

‡ See "Appendix" for description of these statistics.

TABLE VII

Nearest neighbor frequencies of bacterial DNA's

Nearest neighbor sequence	*M. lysodeikticus* DNA (0.41)*	*M. phlei* DNA (0.48)	*A. aerogenes* DNA (0.80)	*E. coli* DNA (1.01)	*B. subtilis* DNA (1.26)	*H. influenzae* DNA (1.62)
ApA, TpT	0.019, 0.017	0.024, 0.026	0.059, 0.061	0.071, 0.076	0.092, 0.095	0.116, 0.116
CpA, TpG	0.052, 0.054	0.063, 0.063	0.067, 0.069	0.071, 0.071	0.067, 0.068	0.067, 0.067
GpA, TpC	0.065, 0.063	0.065, 0.061	0.058, 0.057	0.055, 0.056	0.067, 0.065	0.054, 0.052
CpT, ApG	0.050, 0.049	0.045, 0.045	0.057, 0.056	0.055, 0.055	0.057, 0.058	0.049, 0.050
GpT, ApC	0.056, 0.057	0.060, 0.064	0.052, 0.052	0.055, 0.054	0.048, 0.048	0.048, 0.049
GpG, CpC	0.112, 0.113	0.090, 0.090	0.067, 0.065	0.056, 0.056	0.046, 0.046	0.036, 0.037
TpA	0.011	0.012	0.036	0.051	0.052	0.073
ApT	0.022	0.031	0.053	0.068	0.080	0.095
CpG	0.139	0.139	0.088	0.067	0.050	0.038
GpC	0.121	0.122	0.103	0.083	0.061	0.053
Deviation ratios						
$\sigma_{opp}/\sigma_{error}$	0.6	1.0	0.7	0.8	0.5	0.6
$\sigma_{sim}/\sigma_{error}$	6.7	8.2	6.0	5.3	7.4	7.2
$\sigma_{rand}/\sigma_{error}$	5.8	8.0	6.9	6.1	6.7	7.4

* The numbers in parentheses are the $\frac{AP + Tp}{Gp + Cp}$ ratios for each DNA. (See Table V.)

that the over-all result would show frequencies predicted by random ordering of the nucleotides in the DNA chains. If this were so, the frequency of a given nearest neighbor or dinucleotide sequence (*e.g.* f_{TpA}) in a particular DNA could be predicted from the product of the frequencies of the two constituent mononucleotides (*e.g.* $f_{Tp} \times f_{Ap}$) in that DNA. Thus, the frequencies of sequences ApT and TpA would be equal and predicted by the product $f_{Tp} \times f_{Ap}$. Inasmuch as Ap and Tp occur equally often in DNA, the ApA and ApT sequences would also be equal in frequency. Similarly, the frequencies of the GpA and ApG sequences should on this statistical basis prove to be equal and predicted by the product $f_{Gp} \times f_{Ap}$. The results with all the primers tested (Tables VI, VII, and VIII) show that for every DNA there is significant deviation from the frequencies pre-

TABLE VIII

Nearest neighbor frequencies of bacteriophage DNA's

Nearest neighbor sequence	λ+ DNA (1.00)*	λ dg DNA (1.00)	T2 DNA (1.74)	T4 DNA (1.80)	T6 DNA (1.74)
ApA, TpT	0.069, 0.073	0.072, 0.074	0.111, 0.106	0.109, 0.109	0.106, 0.108
CpA, TpG	0.070, 0.069	0.070, 0.070	0.061, 0.063	0.061, 0.063	0.062, 0.062
GpA, TpC	0.060, 0.064	0.059, 0.061	0.059, 0.057	0.059, 0.059	0.059, 0.058
CpT, ApG	0.056, 0.053	0.056, 0.053	0.054, 0.057	0.056, 0.057	0.058, 0.060
GpT, ApC	0.054, 0.054	0.054, 0.055	0.051, 0.048	0.052, 0.049	0.051, 0.050
GpG, CpC	0.062, 0.063	0.057, 0.056	0.036, 0.034	0.035, 0.032	0.034, 0.033
TpA	0.047	0.047	0.089	0.091	0.090
ApT	0.070	0.068	0.104	0.105	0.101
CpG	0.064	0.069	0.030	0.027	0.028
GpC	0.072	0.079	0.040	0.036	0.040
Deviation ratios					
$\sigma_{opp}/\sigma_{error}$	0.7	0.8	1.1	0.9	0.5
$\sigma_{sim}/\sigma_{error}$	6.3	6.0	5.1	4.8	4.7
$\sigma_{rand}/_{error}$	5.0	5.9	3.9	3.5	3.5

* The numbers in parentheses are the $\frac{Ap + Tp}{Gp + Cp}$ ratios for each DNA. (See Table V.)

TABLE IX

Selective degradation of d-AT copolymer by Lehman E. coli phosphodiesterase

The complete system contained 20 μmoles of glycine buffer (pH 9.7), 2 μmoles of $MgCl_2$, 15 mμmoles of P^{32}-labeled d-AT copolymer (expressed as nucleotide P with a specific activity of 5×10^5 c.p.m. per μmole), 30 mμmoles of enzymatically synthesized "calf thymus" DNA, and 60 units of the *E. coli* phosphodiesterase in a final volume of 0.35 ml. After a 60-minute incubation at 37° the mixture was chilled and 0.75 μmole of native calf thymus DNA was added as "carrier." Immediately, 0.50 ml of cold 1 N perchloric acid was added and the mixture (1.0 ml) allowed to stand at 0° for 5 minutes; it was then centrifuged at $10,000 \times g$ for 5 minutes, and the optical density of the supernatant liquid was determined at 260 mμ. A molar extinction coefficient of 12,000 was assumed for the nucleotides in the acid supernatant liquid. An aliquot of the supernatant liquid (0.5 ml) was also assayed for radioactivity.

System	Amount made acid-soluble	
	P^{32}	Optical density
	%	
1. P^{32}-d-AT (15 mμmoles)	97	97
2. P^{32}-d-AT (15 mμmoles) + DNA (30 mμmoles)	92	34

dicted by random ordering of the mononucleotides. For example, with DNA from *M. phlei* (Table VII), the TpA sequence occurs only half as often as the ApA sequence. With calf thymus DNA (Table VI), the CpG sequence occurs less than a third as often as the GpG sequence. The data of each experiment have been statistically analyzed for deviation from perfect fit with the nearest neighbor frequencies predicted by random arrangement of the mononucleotides (Tables VI, VII, VIII). Some general conformity to arrangements expected from the nucleotide composition is observed and is more marked in certain of the DNA's, for example, in the T-even virulent bacteriophages (Table VIII); yet the sequence frequencies for each DNA sample tested are unique and hardly random in character. The value σ_{rand} expresses the standard deviation of results from those predicted by random arrangement of the nucleotides; it can be seen (Tables VI, VII, and VIII) that these values are consistently higher than the σ_{error} values.

Nearest Neighbor Sequence Frequencies in DNA Synthesized with Native versus Enzymatically Produced Primer—The nearest neighbor sequence frequencies measured by the technique in these studies are those in the newly synthesized DNA. The inference that these frequencies are an accurate reflection of those of the native DNA primer was tested in the following way. An enzymatically synthesized sample of "calf thymus" DNA in which only 5% of the total DNA consisted of the native calf thymus primer was itself used as primer in a typical sequence analysis procedure. The results are compared with those obtained with the native calf thymus DNA as primer (Table VI). It may be seen in Columns 1 and 2 that there is fair agreement between the sequence frequencies of the products primed by the native and by the enzymatically synthesized DNA's, except for notably higher values of TpA and ApT for the latter DNA. This discrepancy was due to a technical difficulty, and the following steps were taken to overcome this.

Reactions on a scale sufficient to produce extensive net synthesis of DNA almost invariably result in some production of the d-AT copolymer (3, 18). Under proper conditions, this polymer represents only a small percentage of the DNA synthesized and does not significantly distort the nucleotide composition of the DNA product (3). However, the d-AT copolymer is 5 to 10 times more efficient as primer than most DNA's and as a consequence is replicated to a significant extent when a DNA sample containing small amounts of the polymer is used for a sequence frequency analysis. More specifically, the sequence patterns are distorted by an abundance of TpA and ApT sequences which comprise this polymer (18 and see below). This interference was overcome by removing contaminating d-AT copolymer from the P^{32}-labeled, enzymatically synthesized DNA before degrading it to the 3′-mononucleotides. This selective removal was accomplished by the use of Lehman's *E. coli* phosphodiesterase (13), an enzyme which degrades heated DNA and single-stranded DNA from bacteriophage ΦX174 (24) but does

not attack native, double-stranded DNA. The data in Table IX show that in a mixture of one part of P^{32}-labeled d-AT copolymer and two parts of unlabeled, enzymatically synthesized DNA, 92% of the d-AT copolymer was degraded to an acid-soluble form whereas less than 5% of the DNA was so affected.

By use of the diesterase to remove the d-AT copolymer contamination (see footnote of Table VI), the sequence frequencies shown in Column 3 of Table VI were obtained. The values for the TpA and ApT sequences are sharply reduced and now agree with those of the native primer; the values for the other sequences are, as a consequence, slightly altered and are in excellent agreement with the native primer.

Nearest Neighbor Sequence Frequencies in Native and Heated DNA—Heating of calf thymus DNA for 30 minutes at 100° in a medium of low ionic strength results in collapse of the rigid helical structure to a randomly coiled configuration. This change in state does not diminish, and in fact improves, the efficiency of the DNA as primer. As shown in Column 4 of Table VI, the sequence frequencies were the same whether native or heated DNA was used as primer.

Nearest Neighbor Sequence Frequencies in DNA of Several Bacteria and Bacteriophages—Data from DNA samples ranging from low to high Ap + Tp content are compared in Tables VII and VIII. Like the data in Table VI, these results agree closely with the base-pairing and opposite polarity of strands demanded by the Watson and Crick model. The sequence frequencies, although showing an over-all influence by the nucleotide composition, are nevertheless not predictable by random arrangement of the nucleotides.

Values for a group of temperate and virulent *E. coli* phages are presented in Table VIII. It may be seen at once that the frequencies for the DNA's of the closely related temperate λ phages and their lysogenic host are very similar. Also, it is apparent that the values for the T-even phages are virtually indistinguishable from one another, as might be expected from their intimate genetic relationship, and further that the values for these virulent phages are strikingly different from the host cell DNA. It should be noted that in the case of the DNA's of the T-even bacteriophages, the data presented in this report are based on the use of deoxycytidine nucleotide, although it is a reasonable presumption that the sequence patterns would be identical if 5-hydroxymethyldeoxycytidine triphosphate were used in place of deoxycytidine triphosphate.

Sequence Frequencies in d-AT Copolymer, d-GC Polymer, and Product of Reaction Primed with Both Polymers—The results in Table X show that chains of the d-AT copolymer contain a perfectly alternating sequence of deoxyadenylate and deoxythymidylate. By contrast, the d-GC polymer consists of unequal amounts of homopolymers of polydeoxyguanylate and of polydeoxycytidylate. When both polymers were present together to prime the enzymatic synthesis, the sequences observed in the product were ApT, TpA, GpG, and CpC, sequences expected from the primers used. These molecules are thus highly ordered structures. Sequences other than those in the original primer, if occurring, are relatively infrequent, and refinements of the current method will be necessary to detect and measure them.

TABLE X

Nearest neighbor frequencies of d-AT copolymer and d-GC polymer

The detailed procedure for the analyses of these polymers was altered in Step A. In the enzymatic synthesis of the P^{32}-labeled DNA (Step A) potassium phosphate buffer (pH 7.4) was substituted for glycine. For the analysis of the d-AT copolymer the synthetic incubation mixture contained no dGTP nor dCTP, and only two reactions were carried out ($dATP^{32}$ + dTTP and dATP + $dTTP^{32}$). Conversely, for the analysis of the d-GC polymer, dATP and dTTP were omitted, and two reactions were carried out ($dGTP^{32}$ + dCTP and dGTP + $dCTP^{32}$). It is known that the omitted nucleotides in each case are not incorporated at all. In the analysis of the d-AT + d-GC mixture the synthetic incubation mixture contained 20 mμmoles of each primer and all four nucleoside triphosphates; all four reactions were carried out in the usual fashion.

In calculating the relative frequencies of nearest neighbor sequences, the computational scheme outlined in the text was applicable only to the analysis of the d-AT copolymer and gave results shown in the table. The general computational scheme could not be used for the data from the analysis of the d-GC polymer or of the d-AT + d-GC mixture; instead, in these cases the amounts of each 3′-nucleotide in the respective digests were determined by dividing its total radioactivity by the specific activity of the originally labeled nucleoside triphosphate. The amounts were then converted to decimal fractions of a unit value.

Nearest neighbor sequence	d-AT copolymer	d-GC* polymer	d-AT + d-GC*
ApA, TpT	<0.001, <0.001		<0.001, <0.001
CpA, TpG			<0.001, <0.001
GpA, TpC			<0.001, <0.001
CpT, ApG			<0.001, <0.001
GpT, ApC			<0.001, <0.001
GpG, CpC		0.525, 0.475	0.170, 0.081
TpA	0.500		0.381
ApT	0.500		0.365
CpG		<0.001	<0.001
GpC		<0.001	<0.001

* Different preparations of the d-GC polymer contain variable and usually unequal amounts of homopolymers of polydeoxyguanylate and polydeoxycytidylate; these differences will be further detailed in a later report.

DISCUSSION

The current results on the nearest neighbor frequencies provide the clearest evidence for a mechanism of enzymatic synthesis of DNA involving the sequential ordering of nucleotides along a DNA template on the basis of hydrogen bonding of adenine to thymine and of guanine to cytosine (25). In addition, these results provide the first proof that the strands of the double helix have the opposite polarity postulated in the Watson and Crick DNA model.

This technique for determining nearest neighbor nucleotide sequences has limitations which warrant discussion. First of all, it cannot, as such, yield information on sequences longer than dinucleotides. The complete degradation of the synthesized DNA to 3′-nucleotides destroys any possibility for probing at trinucleotide or longer sequences. One approach to resolve this problem would be partial rather than total digestion of the synthesized DNA or chemical degradation that would yield 3′-phosphate-terminated dinucleotides or oligonucleotides. In these cases the nucleotide to which the 3′-phosphate end had been attached can be deduced (as was routinely done in these studies). At present there is no way of obtaining a fixed "fingerprint" of

3′-phosphate-terminated oligonucleotides by any degradative method. However, the studies of Rushizky *et al.* (26) with micrococcal DNase and the refined use of acid hydrolysis by Burton and Peterson (6) and by Shapiro and Chargaff (5) offer promising leads in this direction.

Another limitation of the technique is its failure to distinguish sequences involving the uncommon nucleotides, such as 5-methyldeoxycytidylate (MCp) in calf thymus DNA. When calf thymus primer was used, deoxycytidylate always replaced the 5-methyl derivative, which has identical hydrogen-bonding properties, obscuring the important information on the arrangement of 5-methyldeoxycytidine nucleotides and at the same time giving erroneous values for deoxycytidine nucleotide sequences. For example, Sinsheimer's data (4) indicate that the nucleotide sequence MCpG is a major way of positioning 5-methyldeoxycytidylate in calf thymus DNA and could even account for nearly all the sequences registered by our technique as CpG in calf thymus DNA (Table VI). We are hopeful that progress in defining the enzymatic reactions responsible for the presence of the less common nucleotides in DNA will lead to the solution of the sequential arrangement of these nucleotides.

Finally, the limitations in accuracy of these analyses should be pointed out. An analytical error of as little as 1% in our method would permit a relatively large number of errors to be made in determining the sequence frequencies in a DNA molecule. Even the small chromosome of λ phage, containing as it does about 10^5 nearest neighbor sequences, could differ from a related DNA in 1000 nucleotide sequences without any difference being detectable by this technique.

The DNA samples examined in these studies were derived from viral and bacterial sources, calf thymus DNA being the only exception. It would be of considerable interest now to analyze DNA's from a variety of plant and animal sources as well as from a broader selection of bacteria and viruses, including the very interesting single-stranded DNA from bacteriophage ΦX174 (24). From such data, relationships among sequence patterns might come to light that are not obvious to us now. Comparison of the sequence frequency patterns of calf thymus and *B. subtilis* DNA shows striking differences despite their compositional identity. On the other hand, the sequence patterns of DNA's of the temperate viruses λ and λdg are not distinguishable from their host cell, *E. coli*. These observations are thoroughly consistent with the genetic similarities expected of a temperate phage and its host and the differences anticipated between the widely differing bovine and *Bacillus* species.

SUMMARY

1. A technique utilizing 5′-P^{32}-labeled deoxyribonucleotide substrates in the enzymatic synthesis of deoxyribonucleic acid (DNA), followed by enzymatic degradation of the DNA specifically to 3′-deoxyribonucleotides, has yielded the frequency of the 16 possible nearest neighbor arrangements of the 4 deoxyribonucleotides in the DNA chains.

2. DNA samples from viral, bacterial, and animal sources were used as primers, and led to the synthesis of products which in each case had a unique and nonrandom pattern of the 16 nearest neighbor sequence frequencies. In the case of the copolymer of deoxyadenylate and deoxythymidylate and the polymer of polydeoxyguanylate and polydeoxycytidylate, decisive characterization of the primary structure of these DNA's was made possible by this method.

3. The pattern of sequence frequencies was the same in DNA synthesized with calf thymus DNA primer or with enzymatically prepared primer in which only traces of the native calf thymus DNA primer were present. The pattern was also unaltered when the primer used was denatured by heating. These results support the conclusion that the sequence frequencies observed in enzymatically synthesized DNA are the same as in the DNA primer.

4. The distribution of sequence frequencies in 16 DNA samples examined indicates that the mechanism of enzymatic synthesis, and by inference the native DNA structure, is based on base-pairing of adenine to thymine and of guanine to cytosine; the results also furnish experimental evidence for opposite polarity of the two strands of the double helix as suggested by Watson and Crick in their model for DNA.

Acknowledgments—We are indebted to Dr. Lincolin E. Moses for devising the statistical analyses.

We also gratefully acknowledge grants from the Public Health Service which made this work possible.

Appendix—The method of statistical analysis of these data was devised by Dr. Lincoln E. Moses of the Departments of Preventive Medicine and Statistics, Stanford University. He has written the following description of the method. Four ways of gauging the experimental error were applied to the data. Each is appropriate to a different hypothetical model of DNA construction. A close fit of the data to one of the models reflects itself in a small value for the corresponding variance. In each of the four schemes, it was assumed that errors in the percent distribution of counts per minute were of comparable variability, and of true average zero.

The first of the four models is the most general: "The strands of the DNA double helix may be of similar polarity, or they may be of opposite polarity, or conceivably even both polarities occur." Under this model we have, except for experimental error: ApA = TpT and GpG = CpC. The differences in these pairs reflect experimental error, and the sum of squares of these differences (30 in all) divided by 30 gives the estimate σ^2_{error}.

The second model is that with strands only of opposite polarity. In this case, each experiment furnishes six differences for gauging experimental error. They require weighting for the value of the (Ap + Tp)/(Gp + Cp) ratio, and so weighted and appropriately combined lead to the estimate σ^2_{opp}. The calculation is somewhat intricate, and simplified computations were used here. These admit an approximation error of less than 10% in each experiment reported; the error of approximation leads to a slightly high value of σ^2_{opp} in most experiments.

The third model is that with strands only of similar polarity. Here there are eight differences which reflect experimental error; they require weighting for the value of the (Ap + Tp)/(Gp + Cp) ratio. It is readily shown that the eight weighted squared differences provide a six-degree of freedom estimate of experimental error in each experiment. This is the estimate σ^2_{im}.

The fourth model is that of random association of purine and pyrimidine nucleotides in the DNA chains. Under this model the percent distribution of counts for Tp and Ap should be equal to a common value (θ) in each of the four reactions, and the percent distribution of counts per minute for Gp and Cp should be equal to $1 - \theta$ in all four reactions. Thus, there are two sets

of eight quantities hypothetically equal, but for experimental error. The two sums of squares of these quantities about their means, divided by $\frac{44}{3}$, is an unbiased estimate of experimental error variance under this model. The statistic thus computed is σ^2_{rand}.

REFERENCES

1. LEHMAN, I. R., BESSMAN, M. J., SIMMS, E. S., AND KORNBERG, A., *J. Biol. Chem.*, **233**, 163 (1958).
2. BESSMAN, M. J., LEHMAN, I. R., SIMMS, E. S., AND KORNBERG, A., *J. Biol. Chem.*, **233**, 171 (1958).
3. LEHMAN, I. R., ZIMMERMAN, S. B., ADLER, J., BESSMAN, M. J., SIMMS, E. S., AND KORNBERG, A., *Proc. Natl. Acad. Sci. U. S.*, **44**, 1191 (1958).
4. SINSHEIMER, R. L., *J. Biol. Chem.*, **215**, 579 (1955).
5. SHAPIRO, H. S., AND CHARGAFF, E., *Biochim. et Biophys. Acta*, **26**, 608 (1957); **39**, 62 (1960).
6. BURTON, K., AND PETERSON, G. B., *Biochem. J.*, **75**, 17 (1960).
7. WATSON, J. D., AND CRICK, F. H. C., *Nature (London)*, **171**, 737, 964 (1953).
8. SMITH, M., AND KHORANA, H. G., *J. Am. Chem. Soc.*, **80**, 1141 (1958).
9. CUNNINGHAM, L., CATLIN, B. W., AND DE GARILHE, M. P., *J. Am. Chem. Soc.*, **78**, 4642 (1956).
10. HILMOE, R. J., *J. Biol. Chem.*, **235**, 2117 (1960).
11. WITTENBERG, J., AND KORNBERG, A., *J. Biol. Chem.*, **202**, 431 (1953).
12. FISKE, C. H., AND SUBBAROW, Y., *J. Biol. Chem.*, **66**, 375 (1925).
13. LEHMAN, I. R., *J. Biol. Chem.*, **235**, 1479 (1960).
14. SEVAG, M. G., LACKMAN, D. B., AND SMOLENS, J., *J. Biol. Chem.*, **124**, 425 (1938).
15. HERRIOTT, R. M., AND BARLOW, J. L., *J. Gen. Physiol.*, **36**, 17 (1952–1953).
16. KAISER, A. D., AND HOGNESS, D. S., *J. Mol. Biol.*, in press.
17. KAY, E. R. M., SIMMONS, N. S., AND DOUNCE, A. L., *J. Am. Chem. Soc.*, **74**, 1724 (1952).
18. SCHACHMAN, H. K., ADLER, J., RADDING, C. M., LEHMAN, I. R., AND KORNBERG, A., *J. Biol. Chem.*, **235**, 3242 (1960).
19. MARKHAM, R., AND SMITH, J. D., *Biochem. J.*, **52**, 552 (1952).
20. ZAMENHOF, S., BRAWERMAN, G., AND CHARGAFF, E., *Biochim. et Biophys. Acta*, **9**, 402 (1952).
21. LEE, K. Y., WAHL, R., AND BARBU, E., *Ann. Inst. Pasteur*, **91**, 212 (1956).
22. WYATT, G. R., AND COHEN, S. S., *Biochem. J.*, **55**, 774 (1953).
23. CRAMPTON, C. F., LIPSHITZ, R., AND CHARGAFF, E., *J. Biol. Chem.*, **211**, 125 (1954).
24. SINSHEIMER, R. L., *J. Mol. Biol.*, **1**, 43 (1959).
25. KORNBERG, A., *Science*, **131**, 1503 (1960).
26. RUSHIZKY, G. W., KNIGHT, C. A., ROBERTS, W. K., AND DEKKER, C. A., *Biochem. Biophys. Research Communs.*, **2**, 153 (1960).