INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

Xerox University Microfilms
300 North Zeeb Road
Ann Arbor, Michigan 48106

75-28,082

BERMAN, Joy, 1925-
DNA BASE SEQUENCE AND THE CONTROL OF RNA SYNTHESIS: NEAREST-NEIGHBOR ANALYSIS OF RNA TRANSCRIBED IN VITRO FROM NATIVE AND DENATURED BACTERIOPHAGE DNA.

The City University of New York, Ph.D., 1975
Biology

Xerox University Microfilms, Ann Arbor, Michigan 48106

DNA BASE SEQUENCE AND THE CONTROL OF RNA SYNTHESIS: NEAREST-NEIGHBOR ANALYSIS OF RNA TRANSCRIBED IN VITRO FROM NATIVE AND DENATURED BACTERIOPHAGE DNA.

by

JOY BERMAN

A dissertation submitted to the Graduate Faculty
in Biology in partial fulfillment of the requirements
for the degree of Doctor of Philosophy, The City
University of New York.

1975

This manuscript has been read and accepted for the Executive Committee in Biology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

July 2, 1975
Date

Rivka Rudner
Chairman of Examining Committee
Professor R. Rudner

July 7, 1975
Date

Louis G. Moriber
Executive Officer
Professor Louis G. Moriber

S. M. Friedman
Prof. S.M. Friedman

Hunter College
Institution

Ezra Shahn
Prof. E. Shahn

Hunter College
Institution

Burton E. Tropp
Prof. B. Tropp

Queens College
Institution

John Karkas
Prof. J.D. Karkas

Columbia University
Institution

Maurice J. Bessman
Prof. M. Bessman

Johns Hopkins University
Institution

The City University of New York

Abstract

The possible involvement of pyrimidine sequences in the control of DNA transcription has been examined. The frequency of pyrimidine dinucleotides (i.e. TpT, CpC, CpT and TpC) in the DNA template, transcribed *in vitro* by *E. coli* RNA polymerase, has been determined, by inference, from nearest-neighbor analysis of RNA transcripts of T4, T5 and T7 phage DNA. These transcripts were found to be enriched in purine dinucleotides (i.e. ApA, GpG, GpA and ApG) and the DNA template transcribed is therefore enriched in pyrimidine dinucleotides.

For each phage DNA studied the purine enrichment in the native DNA transcripts was determined by: 1. Comparing the dinucleotide frequencies of the RNA transcribed from native and denatured DNA templates; 2. Comparing the frequencies of the matching, complementary dinucleotide pairs, ApA:UpU, GpG:CpC, ApG:CpU, GpA:UpC, ApC:GpU and CpA:UpG; 3. Comparing the dinucleotide frequencies of the RNA products with the values previously determined by nearest-neighbor analysis of the phage DNAs (Josse *et al.*, 1961; Swartz *et al.*, 1962).

Comparing transcripts of native and denatured DNA, increased frequencies were found in: 1. purine dinucleotides ApA, GpA and GpG in the RNA transcripts of native T7 DNA: 2. GpA and ApG in the RNA transcripts of native T4 DNA; 3. ApA and GpA in the RNA transcripts of native T5 DNA. Asymmetry of the RNA transcripts was evaluated from the ratio of the nearest-neighbor frequencies of the complementary dinucleotide pairs. The transcripts of native T7 and T4 DNA were found to be asymmetric. The dinucleotide frequencies of complementary pairs were not equal to each other. In the transcripts of T7 DNA, the increased frequencies of the purine dinucleotides were accompanied

by decreased frequencies of their complementary pyrimidine dinucleotides. In the transcripts of T4 DNA, there was an increase in two purine dinucleotides, and a decrease in their complementary and non-complementary pyrimidine dinucleotides. The transcripts of T5 DNA were less asymmetric than the T4 and T7 RNA products.

The dinucleotide frequencies of the RNA transcripts of both native and denatured DNA templates were found to be non-random. This was determined by comparison with predicted random dinucleotide frequencies, which were calculated from the known base composition of each DNA studied.

It is concluded, from these studies, that short pyrimidine sequences are transcribed by RNA polymerase in control regions of the DNA of phages T4, T5 and T7. Short regular sequences of pyrimidines have been shown to cause alterations of the DNA helix (Langridge, 1969). It is therefore suggested that these pyrimidine sequences, in control regions of the DNA, would constitute such regions of regular sequence, and that they could cause local changes in the structure of the DNA helix. These structural changes can then serve as specific signals for control of DNA transcription. These findings are consistent with the proposal (Chan and Wells, 1974) that local alterations of the DNA helix allow for the formation of unique structures, which can then participate in highly specific protein-nucleic acid interactions, such as the binding of RNA polymerase to promoter sites or the initation of RNA synthesis.

Acknowledgements

I would like to thank Dr. Rivka Rudner for her helpful advice, encouragement and patience, all of which made it possible for me to undertake and to complete this work. I also wish to express my gratitude to Dr. Maurice Bessman, who generously allowed me to complete my experimental work in his laboratory.

In addition, I want to express my appreciation to Dr. Lawrence Viernstein, who very kindly helped me with some of the calculations, and to Dr. Ko-yu Lo for her friendly advice and encouragement.

It is impossible for me to acknowledge the many people who have helped me, in many different ways, in the completion of this work. I am most grateful to all of them.

Table of Contents

List of Tables

List of Figures

Introduction

The transcription of DNA is a carefully controlled process, involving initiation and termination of RNA synthesis at specific sites on the DNA template. This selectivity of transcription is currently believed to be determined, at least in part, by unique DNA sequences. These sequences can be 'recognized' by the DNA-dependent RNA polymerase in a highly specific interaction between the enzyme and DNA. The structural features of such DNA sequences, and the way in which RNA polymerase 'reads' them, is not known. In this study, we wish to examine the proposal that pyrimidine-rich DNA sequences are involved in the control of RNA synthesis (Szybalski *et al.*, 1966).

This proposal is based on two related observations, made in a number of different laboratories. The first is that transcription is an asymmetric process, in which only part of the two strands of the DNA template is transcribed into RNA; this has been shown, in bacterial and bacteriophage systems, *in vivo* (Tocchini-Valentini *et al.*, 1963; Geiduschek and Haselkorn, 1969; Margulies *et al.*, 1970, 1971) and *in vitro* (Geiduschek *et al.*, 1964; Karkas *et al.*, 1968; Summers and Siegel, 1969; Sedat and Sinsheimer, 1970; Pispa and Buchanan, 1971; Brody and Geiduschek, 1970). The second observation is, that in many instances, the DNA strands which are transcribed are pyrimidine-rich, and contain pyrimidine clusters (Marmur and Greenspan, 1963; Summers and Szybalski, 1968; Szybalski *et al.*, 1969; Karkas *et al.*, 1970).

Asymmetric transcription may involve exclusive transcription of only one strand of the DNA duplex, or it may include both DNA strands, switching from one strand to the other, at different times in the cell cycle. For example, RNA is synthesized exclusively from one DNA strand in coliphages T7 and fl, in *B. subtilis* phage SP8 and in phage alpha of *B. megaterium*, where all the *in vivo* RNA transcripts are found to hybridize to only one of the two separated DNA strands (Marmur and Greenspan, 1963; Tocchini-Valentini *et al.*, 1963). On the other hand, in coliphages lambda, T4 and T5, and in animal virus SV40, transcription switches from one strand to the other in going from early to late mRNA synthesis (Cohen and Hurwitz, 1967; Guha and Szybalski, 1968; Lanni and Szybalski, 1969; Sambrook *et al.*, 1972). Asymmetric transcription has also been found in eucaryotic systems. In *Xenopus laevis* and in rat liver, for example, all, or most, of the mitochondrial RNA hybridized with only one strand of the mitochondrial DNA (Dawid and Chase, 1972; Borst and Aaij, 1969).

The chemical asymmetry involved in pyrimidine enrichment of transcribing DNA strands may be limited to an unequal distribution of cytosine and guanine, or it may also include unequal distribution of thymine and cytosine, between the two strands of the duplex DNA (Rudner *et al.*, 1969). The *in vivo* RNA of phages alpha and SP8, and *in vitro* RNA of phage alpha, hybridize exclusively to the DNA strand which is pyrimidine-rich (Tocchini-Valentini *et al.*, 1963; Geiduschek et al., 1964). Each strand of the denatured DNA of these two phages has a different T + C (T: thymine; C: cytosine) content and the strands can therefore be separated in a CsCl gradient, on the basis of their buoyant density (Marmur and Cordes, 1963). In phage lambda, transcription switches from one strand to the other, and the regions transcribed from both strands were found to be pyrimidine-rich (Taylor *et al*., 1967). In the bacteria *E*. *coli* and *B*. *subtilis*, most of the transcription is from the pyrimidine-rich H strand of the DNA (Margulies *et al*., 1970, 1971). In *B*. *subtilis*, all the rRNA and tRNA, as well as 90% of the mRNA, hybridizes to the pyrimidine-rich H strand. The remaining 10% of the mRNA, however, hybridizes to the purine-rich L strand of the bacterial DNA. *In vitro* RNA transcripts of a variety of bacterial DNAs were also shown to be transcribed from pyrimidine-rich DNA; the RNA transcripts were purine-rich and had a base composition that closely resembled a complementary copy of the pyrimidine-rich DNA strands of the bacterial DNA (Karkas *et al*., 1970).

The transcribing strands of DNA, in a number of bacteriophages, were found to contain cytosine-rich and thymine-rich clusters (Szybalski *et al*., 1966). The presence of these clusters was determined by the ability of separated DNA strands to bind to poly (U,G), which binds to cytosine-rich clusters, and to Poly (U), which binds to adenine-rich clusters. Transcribing DNA strands of phages T4, T5, T7 and lambda, for example, were found to bind to poly (U,G) (Guha and Szybalski, 1968; Lanni and Szybalski, 1969; Summers and Szybalski, 1968; Taylor *et al*., 1967). Adenine-rich clusters were found on non-transcribing DNA strands of some phages, such as T4 and lambda, from which the presence of thymine-rich clusters on the complementary transcribing strands was inferred (Guha and Szybalski, 1968; Taylor *et al*., 1967). The exclusive occurrence of these pyrimidine clusters, on the DNA strands found to hybridize with *in vivo* RNAs,led to the proposal that these clusters may play a role in initiation or termination of RNA transcription, or in the selection of the strand to be transcribed (Szybalski *et al*., 1966).

The distribution of pyrimidine clusters (isostichs) has been studied in bacteriophages T7 and lambda (Mushynski and Spencer, 1972a,b), and in the bacteria *E. coli* and *B. subtilis* (Rudner *et al*., 1972; Rudner and LeDoux, 1974). Long pyrimidine isostichs were found to occur predominantly on the transcribing heavy strand of the DNA. In T7 the longest isostichs were found only on the poly (G) binding strand. Twelve components, of chain length 12 or 13, were found in T7, which correlates directly with the same number of *in vivo* mRNA species found by Summers (1969). In Lambda, the longest isostichs are distributed on both strands, as is the poly (G) binding. These poly (G) binding regions on the individual strands of lambda DNA, which contain cytosine-rich clusters, are the regions predominantly transcribed (Taylor *et al*., 1967). In both T7 and lambda phages, long oligocytidylate chains, which would have been expected from poly (G) binding data, were not found, the longest being only 6 or 7 in chain length, respectively. In the analysis of pyrimidine isostichs derived from *E. coli* and *B. subtilis* DNA, an asymmetric distribution of pyrimidine clusters was also found. The asymmetry was most marked in *B. subtilis*, the longest isostichs being found exclusively in the pyrimidine-rich H strand of the DNA. In *E. coli*, the longest pyrimidine oligonucleotides were found predominantly in the H strand. These differences can be correlated with the greater asymmetry shown in the *in vivo* transcription pattern of *B. subtilis*

DNA (Margulies et al., 1970, 1971).

The DNA strands which are preferentially transcribed in RNA synthesis are pyrimidine-rich and contain unique pyrimidine isostichs. It is not known, however, if these pyrimidine regions are actually transcribed. The evidence suggests that they may have a role in transcriptional control, but a role could easily be exclusive of actual transcription into an RNA product. The present studies were concerned with determining whether the RNA asymmetrically synthesized *in vitro* does preferentially incorporate purine sequences. Nearest-neighbor frequency analysis (Josse et al., 1961) was used to determine the dinucleotide frequencies and base compositions of *in vitro* RNA transcripts of DNA from coliphages T4, T5 and T7. By inference, then, the dinucleotide frequencies of the DNA template from which the RNA had been transcribed were also determined (Weiss and Nakamoto, 1961; Hurwitz et al., 1962; Grossman et al., 1971). A brief review of the transcription patterns of these phages will be presented.

In the large virulent coliphage T4 (M.W. 100×10^6, Sinsheimer, 1968) both immediate early and delayed early RNAs hybridize almost exclusively with one strand (1 strand) of T4 DNA (Guha et al., 1971) while late T4 synthesis is principally copied from the other (r strand) (Grasse and Buchanan, 1969). In T4, pyrimidine-rich clusters were found to occur on both strands (Guha and Szybalski, 1968). In the slightly smaller virulent coliphage T5 (83×10^6 daltons, Rubenstein, 1968) the transcriptional pattern, though quite distinctive, has many similarities to that of T4. The DNA of T5 is unique; it consists of one intact and one nicked strand, the single-strand breaks being at genetically determined positions along the molecule. Denatured T5

DNA therefore yields seven or more single-stranded fragments (Bujard and Hendrickson, 1973; Hayward and Smith, 1973). Hybridization of T5 *in vivo* mRNA with the major single-stranded fragments of T5 DNA has shown that both strands are used as template, and that most transcription occurs on the intact heavy strand and on one of the largest fragments of the light strand (Bujard and Hendrickson, 1973; Hayward and Smith, 1973). Hybridization was found to occur preferentially to the single-stranded T5 DNA segments with the most poly (U,G) binding capacity (Lanni and Szybalski, 1969).

In T7,the asymmetry is confined to the exclusive copying of one DNA strand (Summers and Szybalski, 1968). Coliphage T7, smaller than T5 and T4 (26×19^6 daltons, Richardson, 1966), is copied exclusively from the r strand of the viral DNA *in vivo*. *In vivo* RNA of phage T7 was found to hybridize exclusively to the cytosine-rich heavy strand (Summers and Szybalski, 1968). *In vitro* T7 transcription was also found to be exclusively from the heavy strand (Summers and Siegel, 1969).

In the experiments to be described here, *in vitro* RNA transcripts of native and denatured DNA were compared. *E. coli* RNA polymerase, containing the sigma subunit, was used to transcribe T4, T5 and T7 phage DNA. The native DNA of these phages is transcribed asymmetrically by the RNA polymerase holoenzyme (Burgess *et al.*, 1969; Pispa and Buchanan, 1971; Summers and Siegel, 1970). Denatured DNA is transcribed symmetrically, and the high selectivity of the holoenzyme is lost (Chamberlin and Berg, 1964; Brody and Geiduschek, 1970). The dinucleotide frequencies of the RNA transcripts of native and denatured DNA of phages T4, T5 and T7 were determined and compared.

It was assumed that denatured DNA was symmetrically transcribed and that the dinucleotide frequencies of the RNA transcripts were therefore the same as the dinucleotide frequencies of the phage genome. In this way, it was possible to determine if the RNA transcribed from the native DNA template was enriched in purines.

This analysis showed that there is an enrichment of purine sequences in the RNA synthesized on the native DNA, as compared to the RNA transcribed from the denatured template. The DNA template from which the RNA is transcribed is thus not only enriched in pyrimidines, but these pyrimidine-rich regions are transcribed. Specifically, it was found that: 1. Transcripts of native DNA had higher frequencies of purine dinucleotides (i.e. ApA, ApG, GpA and GpG) than the RNA transcribed from the denatured template. In the T7 RNA transcripts, there was an increase in dinucleotides ApA, GpA and GpG. In the T4 RNA transcripts, GpA and ApG were increased, and in the T5 RNA transcripts, there was an increase in dinucleotides ApA and ApG; 2. By evaluating the ratio of the dinucleotide frequencies of the complementary dinucleotide pairs (i.e. ApA:UpU, ApG:CpU, GpA:UpC, GpG:CpC, ApC:GpU and CpA:UpG) it was found that transcription of T7 and T4 native DNA was asymmetric, denatured T7 DNA was transcribed symmetrically, and transcription of native T5 DNA and denatured T4 and T5 DNAs was weakly asymmetric; 3. The dinucleotide frequencies of the RNA transcripts of both native and denatured DNA templates were found to be non-random. This was determined by comparison with predicted random frequencies, which were calculated from the known base composition of each DNA studied.

Materials and Methods

I. Preparation of Escherichia coli RNA polymerase

E. coli RNA polymerase was prepared by the method of Burgess (1969) with some modifications. 100 g frozen *E. coli*/B 3/4 log phase cells were mixed in a Waring blender with glass beads and 100 ml of buffer G (0.05 M Tris-HCl pH 7.5, 0.01 M $MgCl_2$, 0.2 M KCl, 0.1 mM dithiothreitel (DTT), 0.1 mM EDTA, 5% (v/v) glycerol). The cells were blended, with cooling, for 12 minutes. The extract was treated with DNase, washed with 200 ml of buffer G and decanted (Fraction I). Fraction I was then fractionated by ammonium sulfate precipitation. The RNA polymerase was recovered in the fraction precipitated between 33% and 42% saturation. The precipitate was redissolved in 150 ml of buffer A (0.01 M Tris-HCl pH 7.9, 0.01 M $MgCl_2$, 0.1 mM EDTA, 0.1 mM DTT, 5% (v/v) glycerol) (Fraction III) and applied to a DEAE-cellulose column. The column was washed with 50 ml of buffer A and then 400 ml of buffer A + 0.13 M KCl. The RNA polymerase was eluted with 300 ml of buffer A + 0.23 M KCl. The protein concentration of the eluate, as determined by the method of Lowry *et al*., (1951) and the activity of the RNA polymerase (Standard enzyme reaction, Materials and Methods II) are shown in Fig. 1. The peak fractions of the DEAE column, containing the polymerase activity, were pooled (50 ml), reprecipitated with 1.4 volumes of saturated ammonium sulfate, dissolved in a minimal volume (10 ml) of buffer A and dialyzed 3X against 100 volumes of buffer A for 2 1/2 hours at 4°C. Low salt glycerol density gradients (10% - 30% glycerol in buffer A) were run in a Spinco SW 25.1

rotor (29 ml). Twenty five mg of the protein (1.5 ml volume) were applied to each gradient. The gradients were centrifuged at 25,000 rpm for 23 hours at 4°C. Fractions were collected using an auto-Densi Flow Buchler pump, starting from the top of the gradient. The protein pattern of each gradient was followed, as previously, by the O.D. at 280 mu and by Lowry protein determinations. The enzyme activity was assayed by the standard enzyme reaction. Results from one gradient are shown in Fig. 2. The peak polymerase fractions were pooled, and the protein reprecipitated with 1.4 volumes of saturated ammonium sulfate, redissolved in storage buffer (buffer A modified to contain 0.1 M KCl and 50% (v/v) glycerol) and stored at -20°C. The protein concentration of the stored enzyme preparation was 2.6 mg/ml. The steps in the purification procedure are shown in Table 1.

According to Berg _et al_. (1971) poly dAT copolymer is not a suitable template for assays during the early stages of purification, due to its high sensitivity to _E. coli_ nucleases. At later stages, it is the most desirable template since rAU synthesis is largely unaffected by the removal of sigma, and the specific activity obtained for RNA polymerase depends primarily on the activity of the enzyme present. For this reason, T4 DNA was used as a template during the early steps of purification, and dAT was used for all the later stages. The activity of the enzyme is therefore expressed in either T4 or dAT units, where one standard unit of RNA polymerase enzyme activity catalyzes the incorporation of 1 nmole of AMP into a RNA product in 60 minutes at 37°C, using the standard enzyme reaction (Materials and Methods II). In Table 1, therefore, a dAT or T4 unit is equal to one standard enzyme unit of activity, using dAT copolymer or native T4 DNA, respectively, as template. The specific activity of the final fraction was 390 T4 units/mg protein and 840 dAT units/mg protein. This represents approximately a 400 fold purification, based on T4 units. The presence of an active sigma factor was determined by two different methods, as described in Results III.

II. Standard Enzyme Reaction

The standard enzyme reaction was carried out according to Grossman *et al*. (1971), with some modifications. The standard reaction mixture, to measure RNA polymerase enzyme activity, contained the following components, in 250 ul:0.04 M Tris-HCl pH 7.9, 0.01 M $MgCl_2$, 0.1 mM EDTA, 0.1 mM DTT, 0.4 mM phosphate, 0.15 M KCl, labeled ^{14}C or ^{32}P, and unlabeled ribonucleoside triphosphates, (50 nmoles each of ATP, GTP, CTP and UTP), DNA template and RNA polymerase as indicated in the text. Incubation was for 20 minutes at 37°C. At the completion of the reaction, the tubes were placed on ice and 0.1 ml of 0.1 N pyrophosphate, in 0.5 M K_2HPO_4 and 20 ul of bovine serum albumin (5 mg/ml) were added. The product was precipitated with 3 ml of cold 5% trichloroacetic acid (TCA). After 15 minutes at 0° C, the tubes were centrifuged for 5 minutes at 6000 rpm, the supernatant decanted, and the precipitate redissolved in 0.5 ml of 0.1 N NaOH and immediately reprecipitated with another 3 ml of cold 5% TCA. This NaOH and TCA step was carried out at 0°C and completed in about 15 seconds. It was introduced in order to reduce background radioactivity in the TCA precipitate, due to non-specific binding of substrate. The alkali step was not included in the reactions performed for nearest-neighbor analysis. In these assays the TCA precipitate was washed by resuspension in 5% cold TCA. Both of these methods of washing the TCA precipitate reduced the background radioactivity to satisfactorily low levels (approximately 60 cpm as compared to 20 cpm for a filter blank), with no loss of RNA product. After 30 minutes on ice, the TCA precipitate was collected on glass fiber filters (GF/C, Whatman), washed with 5% cold TCA, dried and counted in a Packard Instrument Tri-Carb, by liquid scintillation, using a toluene fluor (toluene containing PPO and dimethyl POPOP).

III. Nearest-neighbor analysis

Nearest-neighbor analysis was carried out according to Grossman *et al.* (1971), with some modifications. Four standard enzyme reaction mixtures, in duplicate, were prepared for each DNA studied. Each reaction mixture contained a different one of the alpha-labeled ^{32}P-ribonucleoside triphosphates and the other three ribonucleoside triphosphates, unlabeled. (Specific activity of the labeled substrates was 2 to 13 x 10^3 cpm/nmole.) Amounts of DNA template and RNA polymerase are described in Results I, Table 2. The set of eight reaction tubes (two tubes for each of the four labeled substrates) were incubated, simultaneously, for 20 minutes at 37°C. All results reported are average values for the duplicate reactions. Each reaction mixture was precipitated and washed 1X with 5% cold TCA, as described above in the standard enzyme reaction. The washed TCA precipitate was dissolved in 1 ml of 0.3 M KOH, and hydrolyzed, for 18 hours at 37°C, to 2'-(3')-nucleoside monophosphates. Carrier

2'-(3')-nucleoside monophosphates (13-20 nmoles of each of the four bases, in 5 ul of water) was added to each hydrolysate. The 1 ml hydrolysate was then desalted on small Dowex - 50/H+ columns, by elution with 10 ml of water. The samples were evaporated to dryness and were redissolved in 150-200 ul of water. Two 75-100 ul portions of each sample were spotted for two parallel electrophoretic separations on Whatman 3MM paper. Paper electrophoresis was carried out in citrate buffer (0.025 M pH 3.5) at 1500 V for 90 minutes in an electrophoresis cell, using CCl_4 as coolant. The papers were oven-dried and the UV absorbant nucleoside monophosphates identified by examination under ultraviolet light. The four ribonucleotide spots were cut from the paper and the ^{32}P radioactivity determined.

IV. Specific activity of labeled ribonucleoside triphosphates

The specific activity of the 8-14C and alpha-labeled ^{32}P substrates was determined by paper electrophoresis. Each triphosphate was spotted on Whatman 3MM paper and electrophoresis was carried out in citrate buffer (0.025 M, pH 4,9) at 1500 V for 90 minutes. The nucleoside triphosphate spot was identified under ultraviolet light, cut and eluted in 3 ml of 0.05 N HCl. The molar concentration of each triphosphate was determined by its optical density at peak wavelength (pH 2.0) and the radioactivity was determined. The specific activity was approximately 1 x 10^3 cpm/nmole for the ^{14}C labeled substrates and from 2 to 13 x 10^3 cpm/nmole for the ^{32}P labeled substrates.

V. Analysis of transcription products

The RNA transcripts synthesized for nearest-neighbor analysis were analyzed by nuclease digestion, to determine the amount of single-stranded RNA, RNA:DNA hybrid or double-stranded RNA synthesized in each reaction. The radioactive product remaining after digestion was determined by acid precipitation and nitrocellulose membrane filtration. ^{14}C transcripts were made with each DNA template, using the standard enzyme reaction (reaction volume:500 ul). The DNA template and RNA polymerase were at the same ratios and concentrations as were used in the synthesis of the ^{32}P-RNA for nearest-neighbor analysis (Table 2). The first part of the reaction, i.e. incubation for 20 minutes at 37°C, was carried out in a single reaction mixture. Following the initial incubation the sample was placed on ice and immediately diluted with an equal volume of diluent containing 0.1 M NaCl and 0.01 M $MgCl_2$ (Geiduschek and Grau, 1970). The final concentration of the buffer was then: 0.02 M Tris-HCl pH 7.9, 0.05 mM DTT, 0.05 mM EDTA, 0.20 mM PO_4, 0.75 m KCl and 0.05 M NaCl and 0.01 M $MgCl_2$ (Geiduschek and Grau, 1970). 100 ul aliquots of the reaction mixture were removed, and one of the following additions was made: RNase T1 (1 ug/ml) and RNase A (2 ug/ml), DNase (30 ug/ml) or both RNase and DNase (Geiduschek and Grau, 1970; Robertson, 1971). All samples were incubated for an additional 30 minutes at 37° C. Under these conditions, DNase digestion made *E. coli* native ^{32}P-DNA 75% acid soluble, and RNase digestion (RNase A and RNase T1) made ^{3}H-RNA of *B. subtilis* 100% acid soluble. After the second incubation, two aliquots (45 ul each) were removed from each sample and treated as follows: one aliquot was diluted into 10 ml of 2X standard saline-citrate (SSC, standard, 0.15 M NaCl, 0.015 N Na Citrate pH 7.0), filtered on a nitrocellulose membrane filter (Schleicher and Schuell Co., B6, 24 mm) using a Millipore filter apparatus, and washed with 30 ml of 2X SSC; the other aliquot was precipitated with 5% cold TCA. All samples were dried and the radioactivity determined.

VI. Preparation and denaturation of bacteriophage DNA

DNA was isolated from phage T4, according to the procedure of Grossman *et al*. (1961). The phage was extracted three times with a 0.01 M potassium phosphate buffer (pH 7.0) - saturated phenol mixture. The phenol was removed from the aqueous DNA layer by ether extraction. The DNA was precipitated by addition of cold ethanol, and dissolved in 1X SSC. DNA of phages T5 and T7, prepared by phenol extraction, were kindly supplied by Dr. M. J. Bessman. Double-stranded RF DNA of phage f1 was kindly supplied by Dr. P. Modell.

All phage DNAs were denatured by alkali. The DNA (3-5 ml volume), at a concentration of 80-100 ug/ml, in 0.1X SSC, was brought to a pH of 12.5 by addition of 0.1 volume of 3 N NaOH. After stirring for 10 minutes at room temperature, the solution was neutralized with 3 N HCl and then dialyzed against 1X SSC for 18 hours at 4°C.

VII. Materials

Superbrite 100 glass beads were purchased from 3M Company. *E. coli*, 3/4 log phase frozen cells, grown on enriched medium at 37°C, were purchased from Grain Processing Corporation. Whatman microgranular diethylamine cellulose (DEAE) DE-52 (1.0 mg/gm dry weight) was purchased from Reeve Angel Company. Enzyme grade ammonium sulfate was purchased from Mann Research Laboratories.

^{14}C labeled ribonucleoside triphosphates, unlabeled ribonucleoside triphosphates and 2'-(3')-nucleoside monophosphates were purchased from Schwartz Bioresearch. Alpha-^{32}P labeled ribonucleoside triphosphates were purchased from New England Nuclear.

Copolymer poly (dA-dT) was purchased from PL Biochemicals, Inc. *E. coli* ^{32}P-DNA was kindly supplied by Dr. M. J. Bessman. *B. subtilis* ^{3}H-RNA was kindly prepared by Dr. Y. Setoguchi, according to Margulies *et al.*, (1970). DNase I, RNase A and RNase T1 were purchased from Worthington Biochemical Company.

Results

I. Determination of optimum reaction conditions for transcription of native and denatured phage DNA

In these studies, the dinucleotide frequencies of RNA transcribed asymmetrically from native phage DNA were compared with the dinucleotide frequencies of RNA transcribed from denatured phage DNA. The optimum reaction conditions for these two types of RNA synthesis, for phages T4, T5 and T7, had to be determined for each phage studied. The optimum conditions for each type of synthesis, as will be explained below, depends on the DNA:polymerase ratio at which the transcripts are made.

In vitro transcription of phages T4, T5 and T7 native DNA, with *E. coli* RNA polymerase, is an asymmetric process and is initiated at specific sites on the genome. The *E. coli* RNA polymerase must contain the sigma subunit and the template must be native DNA, for this specificity to be retained *in vitro* (Burgess *et al.*, 1969; Hinkle and Chamberlin, 1972; Pispa and Buchanan, 1971). At low molar ratios of DNA to enzyme it is possible, however, that some initiation may occur at other than true promoter sites (Dause *et al.*, 1972). It is assumed, therefore, that at high DNA to enzyme ratios, all of the RNA polymerase binds to true promoter sites and the selectivity of transcription is retained. All RNA transcripts of native DNA were therefore synthesized at high DNA to enzyme ratios, at which the DNA was present in saturating amounts.

In vitro transcription of denatured DNA of phages T4, T5 and T7, yields a different type of RNA product than the one just described. In the denatured DNA there is an alteration of the DNA template structure and a loss of much of the specificity of initiation. It is possible, in this case, for almost all of the DNA to be transcribed (Chamberlin and Berg, 1964). For the purposes of these studies, the RNA transcripts of the denatured template are considered to be approximate copies of the DNA duplex. The denatured DNA template, however, does retain some degree of specificity, as has been shown with phage T4 DNA (Brody and Geiduschek, 1970). It is assumed, however, that the initiation of RNA synthesis at possible preferred sites, and transcription of selective regions of the DNA, will be minimized if there is an excess of polymerase available. Therefore, all the RNA transcripts of denatured DNA were made at low template to polymerase ratios, at which the enzyme was present in saturating amounts.

Saturating amounts of DNA or enzyme were determined for each native or denatured DNA template, respectively. Saturation curves for native and denatured T5 and T7 DNA templates are shown in Figure 3. Similar curves were obtained for T4 DNA. RNA synthesis was measured at varying DNA:enzyme ratios, in the standard enzyme reaction (Materials and Methods II). At constant enzyme concentration, such titration curves have an initial linear portion and ultimately reach a plateau of synthesis, insensitive to further addition of DNA. In these studies, enzyme saturation is defined by the linear portion of each curve, and DNA saturation by the plateau. The optimum DNA:protein ratio chosen for transcription of each type of DNA is shown (Fig. 3) by the arrow. DNA saturation was achieved at a DNA:protein ratio of approximately 0.5 for the native DNA templates (Table 2). Native T7 and T5 DNA were transcribed at this minimum ratio. Native T4 DNA was transcribed at the higher ratio of 1.0 (see Legend, Table 2, for explanation). Denatured DNA was found to be a less efficient template for RNA synthesis than native DNA. To compensate for this, the maximum DNA:protein ratio, 0.15 to 0.20, on the linear portion of the curve, was chosen for RNA synthesis with the denatured DNA templates. The DNA:protein ratios, determined from the saturation curves, and the amount of DNA template and RNA polymerase protein used in the present studies are summarized in Table 2. The amount of RNA synthesized in each reaction, and the percentage of the available template transcribed by the RNA polymerase, are also shown.

The several DNA templates varied considerably in efficiency (Table 2). The template efficiency (nmoles RNA) synthesized/ug protein) of the native DNAs was consistently higher than that of the denatured DNAs. Native T7 DNA had the highest efficiency (1.4 nmoles RNA synthesized/ug protein). Native T4 and T5 DNA had lower efficiencies (0.87 and 0.82 nmoles RNA synthesized/ug protein). The denatured DNAs all had approximately one-tenth the template efficiencies of their respective native DNAs, at the chosen ratios. The efficiencies of the native and denatured templates are not strictly comparable since they have been determined under two different saturation conditions. In Figure 3, the native template can also be seen to be more efficient, when synthesis of both types of template is compared at DNA saturation levels.

These results are in agreement with previous findings that denatured DNA is a much poorer primer for RNA synthesis than native DNA (Maitra and Hurwitz, 1965; Salvo _et_ _al_., 1973). The basis for

√ this difference in priming efficiency of native and denatured templates, with E. coli RNA polymerase, is not known. It may be due to a reduced rate of synthesis, premature termination or loss of in vivo initiation sites, which require the double-stranded DNA structure. The E. coli RNA polymerase may also bind to the denatured DNA template in several different ways, and the enzyme may not necessarily be able to initiate synthesis at all of these sites (Hinkle et al., 1972).

The amount of DNA and RNA polymerase used for each experiment, at the chosen DNA:polymerase ratios, was selected with the purpose of normalizing the amount of product made in each reaction. In order to compensate for the lower efficiency of the denatured DNA templates, the amount of enzyme used to transcribe the denatured DNA was always much greater than the amount used with the native DNA templates. The one exception to this was the experiment involving limited synthesis with native T4 DNA, where the reaction mixture was incubated for 6 minutes at 23° C. For an explanation and discussion of this experiment see Results IV.

II. Analysis of products formed in transcription of native and denatured DNA templates

The RNA transcripts of T4, T5 and T7 phage DNA, synthesized in these studies, were analyzed to determine the specific RNA products formed in our reaction system. The RNA products of *in vitro* transcription of native and denatured DNA include single-stranded RNA, DNA:RNA hybrids and possibly, a double-stranded RNA (Geiduschek *et al.*, 1961; Warner *et al.*, 1963; Bishop, 1969; Robertson, 1971). Transcription *in vitro* of native DNA produces a predominantly single-stranded RNA product, readily separable from its DNA template in a CsCl density gradient (Geiduschek *et al.*, 1961). This RNA product is also very sensitive to RNase digestion. For example, less than 5% of the RNA transcripts of native T4 DNA are RNase resistant (Bishop, 1969). In contrast, transcription *in vitro* of single-stranded DNA produces an RNase resistant DNA:RNA hybrid (Warner *et al.*, 1963). The formation of DNA:RNA hybrids has been shown, with such templates as denatured T4 DNA (Bishop, 1969), denatured calf thymus DNA (Warner *et al.*, 1963) and the single-stranded DNA of bacteriophage ØX 174 (Chamberlin and Berg, 1964). A double-stranded RNA product was also reported with denatured T4 DNA (Bishop, 1969) and single-stranded f1 DNA (Robertson, 1971) as templates for *in vitro* RNA synthesis. We wished to establish whether similar products, especially RNA:RNA hybrids, were synthesized in our system.

The RNA transcripts were analyzed by their susceptibility to digestion by RNase, DNase or a combination of both enzymes. The amount of single-stranded RNA, DNA:RNA hybrid or double-stranded

RNA in each digest, was determined by the proportion of acid precipitable RNA that was found to bind to nitrocellulose membranes. Briefly, the types of RNA products present were determined as follows: 1. The membrane bound RNA product is DNA:RNA hybrid ; 2. The amount of acid precipitable RNA rendered acid soluble, after RNase digestion, is single-stranded RNA; 3. Double-stranded RNA is the only product remaining as non-membrane bound, acid precipitable RNA after combined DNase and RNase digestion. The results of these experiments (Table 3) showed differences in the RNA products formed, in transcription of native and denatured DNA templates, that were consistent, in part, with the published findings already cited. The transcripts of native DNA were predominantely single-stranded RNA. This was shown as follows: 1. 85-90% of the acid precipitable RNA product was not retained by membrane filtration; 2. After RNase digestion, only 2-5% of the acid precipitable radioactivity remained; 3. The small amount of RNA remaining after RNase digestion was also membrane bound and therefore, DNA:RNA hybrid. The reduction of RNA product retained by the membrane, i.e., loss of presumed DNA:RNA hybrid, after RNase digestion could have been due to digestion of loose ends of RNA in the hybrid. It could also have been due to loss of single-stranded RNA, that was membrane bound by the enzyme, in an RNA-DNA-polymerase complex. In the latter case, more than 95% of the RNA transcribed from native DNA would be single-stranded RNA.

The transcripts of denatured DNA had a large proportion of DNA:RNA hybrid product. This was shown as follows: 1. 90% (T4 DNA) and 30% (T7 DNA) of the acid precipitable RNA products were membrane-bound; 2. After RNase digestion, as much as 30% of the acid precipitable product was retained by the membrane. In contrast, only 2-5% of the native DNA transcripts were retained after RNase digestion; 3. All of the denatured T4 DNA transcripts that were acid precipitable after RNase digestion (i.e., 30%), were also retained by membrane filtration. They were therefore predominantly DNA:RNA hybrid; 4. There was a reduction in acid precipitable RNA product after RNase digestion, suggesting that a significant proportion of membrane-bound product may be single-stranded RNA, bound by the enzyme, in an RNA-DNA-polymerase complex.

Digestion of the reaction products with DNase confirmed the results found with RNase. After DNase digestion, the transcripts of native DNA showed a greatly reduced amount of membrane-bound DNA:RNA hybrid. Only about 1-5% of the acid precipitable RNA was still membrane-bound. This suggests that most of the RNA, previously bound to the membrane, may have been single-stranded RNA, bound in an RNA-DNA-polymerase complex. The DNase digests of the transcripts of denatured DNA showed a significantly higher proportion of membrane-bound DNA:RNA hybrid (20-30%). Combined digestion with DNase and RNase, indicated that only 2-3% of the RNA transcripts of the denatured T4 and T7 DNAs could be double-stranded RNA. The difference between the acid precipitable and membrane-bound RNA, for these denatured DNAs, i.e., 2.4% for T4 and 1.8% for T7, could be double-stranded RNA.

In conclusion, based on this analysis, the transcription products used here for nearest-neighbor analysis, were single-stranded RNA or DNA:RNA hybrid, as was described by others (Geiduschek _et_ _al_., 1961; Chamberlin and Berg, 1964). There may be a very small amount of double-stranded RNA present in the denatured DNA transcripts, but this small amount should not affect the results obtained in the nearest-neighbor analysis.

III. The selectivity of in vitro RNA synthesis: the presence of sigma

E. coli RNA polymerase must contain an active sigma subunit in order to correctly initiate selective transcription on the native DNA template. Two experiments were carried out, in order to establish that the RNA polymerase, prepared for these studies, had an active sigma factor. In the first experiment, the base composition of the RNA product of transcription of f1 replicative form (RF) DNA was determined. In the second experiment, the efficiency of the enzyme, in transcribing native T4 DNA and copolymer dAT, was compared.

The double-strained RF form of f1 DNA presents an ideal template for determining the specificity of transcription of *E. coli* RNA polymerase. Correct transcription is confined exclusively to one strand of the DNA and the base composition of the product is known. The minus strand of the f1 RF DNA is the sole template for *in vivo* transcription (Sedat and Sinsheimer, 1970). The *E. coli* RNA polymerase holoenzyme, containing the sigma subunit, also transcribes only the minus strand *in vitro*, whereas both strands are transcribed by the core enzyme, which lacks sigma (Suguira *et al.*, 1970). The largest transcripts made by the holoenzyme, 26S, correspond roughly to the length of the f1 genome (Okamoto *et al.*, 1969). The base composition of correctly transcribed *in vitro* RNA is complementary to the minus strand and identical to that of the plus strand, whose composition is: A: 0.245; G: 0.205; C: 0.202; T: 0.348 (Schaller *et al.*, 1969). The composition of the RNA product therefore provides an excellent means for determining if the *E. coli* RNA polymerase has the required specificity.

An independent method for the determination of the base composition of an RNA product is provided in the method of nearest-neighbor analysis (Josse et al., 1961). It was therefore possible to determine the base composition of the RNA product of f1 transcription by the same method as was to be used for the dinucleotide frequency analysis of the RNA transcripts of T4, T5 and T7 phage DNA. Nearest-neighbor analysis was carried out on in vitro transcripts of f1 RF DNA (for experimental details see Materials and Methods, III). The ^{32}P content and fraction of the total radioactivity transferred to each of the four 2'-(3') ribonucleoside monophosphates were determined for each substrate and the values are given in Table 4A. The nearest-neighbor frequencies of the dinucleotide pairs and the base composition of the RNA product were then calculated, according to the method outlined in Appendix A. The base composition of the RNA product (Table 4B) was found to be the same as that of the f1 plus DNA strand (A = 0.242, G = 0.202, C = 0.205, U = 0.346), showing that the RNA polymerase was selectively copying only the minus strand of the template. The dinucleotide frequencies calculated (Table 4C) will be discussed below (Results IV) with the T phage results.

Using another method, the presence of sigma in the E. coli RNA polymerase was established by the relative efficiency of the enzyme in transcribing native T4 DNA and dAT copolymer. T4 native DNA is a very efficient template for the sigma containing holoenzyme, but a very poor one for the core enzyme (Burgess et al., 1969). Core enzyme makes only 3-4% of the amount of RNA product synthesized with the holoenzyme. For poly dAT, however, synthesis is largely unaffected by the removal of sigma (Berg *et al.*, 1971). The dAT copolymer is approximately three times more efficient than the T4 DNA, as a template for RNA synthesis with the *E. coli* holoenzyme, but approximately fifty times more efficient with the core enzyme. The relative amount of RNA made on these two templates, therefore, indicates whether or not the RNA polymerase being used contains sigma. The specific activity of the polymerase with each template was determined, using the standard enzyme reaction (Materials and Methods, II). The specific activity of the enzyme was 390 units/mg protein with the native T4 DNA template and 840 units/mg protein with the dAT template. (The unit of enzyme activity, used here, is defined in Materials and Methods I). The amount of product made with the dAT template was therefore approximately twice as much as with the T4 DNA. This is the ratio that would be expected for transcription with *E. coli* RNA polymerase containing the sigma subunit.

The *E. coli* RNA polymerase prepared, as described in Materials and Methods I, should contain an active sigma subunit. (There is no step in the procedure which is believed to dissociate that subunit from the rest of the enzyme.) The two experiments described here confirm the presence of an active sigma subunit in the enzyme.

IV. Nearest-neighbor analysis of in vitro RNA transcripts of native and denatured phage DNA

Nearest-neighbor analysis was carried out on *in vitro* RNA transcripts of native and denatured DNA from bacteriophages T4, T5 and T7. The *in vitro* RNA product provides nearest-neighbor data for its DNA template (Weiss and Nakamoto, 1961; Hurwitz *et al.*, 1962; Grossman *et al.*, 1971). As was discussed in the introduction, comparison of the two types of RNA transcript provides a basis for determining if an enrichment of specific dinucleotides exists in the DNA template which is transcribed by RNA polymerase.

The nearest-neighbor frequencies of the sixteen possible dinucleotide sequences of the RNA product were determined according to Josse *et al.*, (1961) (Materials and Methods, II). Radioactivity measurements of ^{32}P-label transferred from 5'-nucleotides to 2'-(3')-mononucleotides, for the RNA transcripts of T7, T4, T4-limited and T5 DNAs, are given in Tables 5A, 6A, 7A and 8A, respectively. The T4-limited experiment, which restricted synthesis to immediate early message, is discussed fully later in this section. The calculation of the dinucleotide frequencies of the RNA product is described in Appendix A.

The asymmetry ratio of the RNA transcripts. In the double-stranded DNA structure, established by nearest-neighbor analysis of enzymatically synthesized ^{32}P-DNA, there are six pairs of matched, complementary dinucleotides that have equivalent frequencies: ApA=TpT, GpA=TpC, ApG=CpT, GpG=CpC, CpA=TpG and GpT=ApC. The ratio of the nearest-neighbor frequencies of these matching dinucleotides is, therefore, equal to 1.0. This ratio can be used to determine how closely the nearest-neighbor frequencies of the RNA transcripts resemble the nearest-neighbor frequencies of the DNA genome. The ratio of the nearest-neighbor frequencies of a matching pair of dinucleotides is defined, for these studies, as the asymmetry ratio. The asymmetry ratio, of any one matching dinucleotide pair, may be: 1. Equal to 1.0; transcription of the pair is symmetric; 2. Not equal to 1.0: transcription of the pair is asymmetric. In the latter case, the asymmetry may involve: 1. An enrichment of purine dinucleotide frequencies, e.g. ApA greater than UpU; 2. An enrichment of pyrimidine dinucleotide frequencies, e.g. CpU greater than ApG; 3. An enrichment in purines and pyrimidines, e.g. ApC greater than GpU. It is also possible for asymmetric transcription to occur in such a way as to maintain the equality of these paired dinucleotides, and an asymmetry ratio of 1.0. This could happen, for example, in the transcription of palindromes. Palindromes contain complementary sequences which are arranged on single-stranded DNA in such a way, that self-folding of the DNA could form a double-stranded DNA, leaving out only a few unmatched bases (Sekiya and Khorana, 1974; Maniatis and Ptashne, 1974; Wilson and Thomas, 1974).

The base composition and base ratios of the RNA transcripts. Chemical asymmetry in the transcribed DNA template may be inferred from the base composition and base ratios of the RNA transcripts. The base composition of each RNA transcript was determined, as described in Appendix A, from the nearest-neighbor data, according to Josse *et al.* (1961). The base compositions and base ratios for the RNA transcripts of T7, T4, T4-limited and T5 DNAs are given in Tables 5B, 6B, 7B and 8B, respectively (selected values shown below). As described in the introduction, transcribing DNA strands are pyrimidine-rich (Marmur and Greenspan, 1963; Szybalski et al., 1969; Mushynski and Spencer, 1970a,b). RNA transcripts of these strands would be purine-rich if the pyrimidine-rich regions are transcribed. In RNA transcripts, in which both strands of the DNA are copied symmetrically, the base complementarity of the duplex DNA would be maintained in the RNA product. In these studies, the base composition of the RNA transcripts of native DNA was found to be enriched in purines. The T7 DNA transcripts were high in A. The T4, T4-limited and T5 DNA transcripts were high in G.

Selected values from Tables 5B, 6B, 7B and 8B

Phage	Base or Ratio	Mole % or Ratio	
		Native DNA	Denatured DNA
T7	A:	0.279	0.262
	A/U:	1.13	1.02
T4	G:	0.181	0.188
	C:	0.161	0.157
	G/C:	1.13	1.20
T4-limited	G:	0.185	
	C:	0.167	
	G/C:	1.11	
T5	G:	0.203	0.225
	C:	0.188	0.183
	G/C:	1.08	1.23

The mole fraction of A, in the native T7 RNA transcripts, was higher than the amount of A in the denatured T7 RNA transcripts, and the amount of T in T7 phage DNA. The A/U ratio of the native T7 RNA transcripts was 1.13 and the other base ratios (i.e., G/C, purine/pyrimidine and 6-amino/6-keto) were also greater than 1.0. The base ratios of the denatured T7 RNA transcripts, however, were all close to 1.0. These values show a chemical asymmetry, and purine enrichment, in the RNA transcripts of the native T7 DNA. This asymmetry was not present in the denatured T7 RNA transcripts. In the T4 and T4-limited RNA transcripts, there was a relative rather than an absolute increase in G, since the mole fraction of G was essentially unchanged, while the mole fraction of C decreased. In the T5 RNA transcripts, there was some increase in G and also a decrease in C. These values were compared to their complementary bases in the phage DNAs. The purine/pyrimidine ratio remained close to 1.0 in this group of RNA transcripts. In contrast to the results with the T7 DNA, where the base composition and base ratios of native and denatured RNA transcripts followed a pattern of asymmetric and symmetric transcription, the values for the T4 and T5 RNA transcripts suggest that both the native and denatured DNA templates were transcribed asymmetrically. The G/C values were greater than 1.0 for both types of transcripts. The other base ratios were similar to each other in both the native and denatured DNA RNA transcripts.

Variance of the asymmetry ratio. The calculated nearest-neighbor frequencies (NNF) of the *in vitro* RNA transcripts of T7, T4, T4-limited and T5 phage DNA, shown in Tables 5C, 6C, 7C and 8C, respectively, have been arranged so that the first four sequences in column NNF I are all purine dinucleotides (i.e. ApA, ApG, GpA and GpG) and the matching sequences, in column NNF II, are pyrimidine dinucleotides (i.e., UpU, CpU, UpC and CpC). An asymmetry ratio significantly greater than 1.0, for these four dinucleotide pairs, indicates an enrichment of purine sequences in the RNA product. An asymmetry ratio significantly less than 1.0 indicates a decrease of purine sequences in the RNA product. The other two matching pairs are mixed purine-pyrimidine dinucleotides (i.e. CpA:UpG and GpU:ApC). Asymmetry ratios significantly different from 1.0, for these two pairs, indicate an enrichment of purines A or G, not purine sequences, in the RNA product.

The nearest-neighbor frequencies of the six matched dinucleotide pairs were evaluated for equivalence by the 't' test and the confidence level, p, determined. In this analysis, the asymmetry ratio was considered to be significantly different from 1.0, and the frequencies NNF I and NNF II were determined to be unequal, when p was less than or equal to 0.05. The asymmetry ratio was not considered to be significantly different from 1.0, and the frequencies NNF I and NNF II were considered to be equal, when p was greater than 0.05.

The degree of asymmetry of transcription of each DNA template was estimated from the variance* of each set of six asymmetry ratios. The variance is a measure of the extent to which each asymmetry ratio differs from the predicted value (1.0) for symmetric transcription. Based on the variance of the asymmetry ratios, the degree of asymmetric

---

* See Appendix B for an explanation of the calculation of the variance.

transcription of native and denatured templates was compared (Table 9). The variance is calculated from the deviation of each asymmetry ratio from 1.0 (variance observed). The expected variance of the predicted value (1.0), due to experimental error, must also be determined (variance error). The ratio of these two variances, i.e. variance observed/variance error, and the number of degrees of freedom involved, determines the level of probability, P, at which the observed variance is significantly greater than experimental error, and is thereby different from the predicted value (Table for F test for equality of variances: Snedocor). For values of P greater than 0.05, the variance of the asymmetry ratios was taken to be equal to the experimental variance; the asymmetry ratios were considered to the equal to 1.0, and transcription assumed to be symmetric. For values of P less than 0.05, the observed variance was taken to be significantly different from the experimental variance; the asymmetry ratios were considered to be different from 1.0, and transcription assumed to be asymmetric.

Analysis of nearest-neighbor frequencies of the RNA transcripts. The asymmetry ratios of the six matching complementary dinucleotide pairs in the RNA transcripts of T7 DNA were analyzed (Table 5C, selected values shown below). In the transcripts of native DNA, an enrichment of purine sequences was found for three of the first four pairs; the asymmetry ratios were significantly greater than 1.0. The ratios were significantly less than 1.0 for two of the pairs, and in the sixth pair the frequencies were equal. In the RNA transcripts of denatured T7 DNA, asymmetry ratios in four of the six pairs were not significantly different from 1.0.

Selected values from Table 5C.

Dinucleotide pair	Asymmetry Ratios Native	(T7 DNA template) p	Denatured	p
ApA:UpU	1.18	0.001	1.02	0.20
GpA:UpC	1.26	0.001	1.03	0.10
ApG:CpU	0.82	0.001	0.92	0.001
GpG:CpC	1.32	0.001	0.96	0.20
CpA:UpG	1.02	0.30	1.01	0.50
GpU:ApC	0.80	0.001	0.90	0.001

The variance of the asymmetry ratios for the native T7 RNA transcripts (Table 9), for four and six dinucleotide pairs (7.8 and 5.5, respectively), was found to be significantly greater than the variance calculated for the experimental error (0.09 and 0.08, respectively), and the asymmetry ratios were determined to be significantly different from 1.0. For the denatured T7 RNA transcripts, the variances observed were much smaller (0.31 and 0.39) and the asymmetry ratios were determined to be not significantly different from 1.0. Following the criteria proposed here, transcription of native T7 DNA was clearly asymmetric, while transcription of denatured T7 DNA was symmetric. The asymmetry ratios and the variance show that there was an enrichment of purine sequences in the RNA transcripts of native T7 DNA. This finding is consistent with the observation that all T7 _in vivo_ and _in vitro_ transcription occurs on the r strand (Summers and Szybalski, 1968; Summers and Siegel, 1969), which also contains all the long pyrimidine oligonucleotides that have been found in T7 DNA (Mushynski and Spencer, 1970a). The purine enrichment, found in the RNA transcripts in the present studies, suggests that pyrimidine-rich regions on the r strand of T7 DNA are transcribed into an RNA product.

The base composition of the RNA transcripts of T7 native DNA was high in A, but not in G (Table 5B). However, when the individual dinucleotide frequencies of the native and denatured T7 RNA transcripts were compared (Table 5D, selected values shown below), the frequencies of six dinucleotides, containing both A and G, were higher in the transcripts of native DNA.

Selected values from Table 5D.

Dinucleotide Sequence	Frequency (RNA)		Base change (in Native) + increase - decrease 0 no change
	Native	T7 DNA template Denatured	
ApA	0.0804	0.0668	+ A
GpA	0.0734	0.0683	+ G,A
UpC	0.0583	0.0664	- U,C
ApG	0.0599	0.0671	- A,G
GpG	0.0545	0.0510	+ G
CpC	0.0414	0.0532	- C
GpU	0.0590	0.0627	- G.U
ApC	0.0728	0.0696	+ A,C
ApU	0.0588	0.0575	+ A,U
GpC	0.0682	0.0556	+ G,C

In the native T7 RNA transcripts, the frequencies of three purine dinucleotides increased. For two of these, the frequencies of their complementary pyrimidine dinucleotides decreased, thereby increasing the asymmetric transcription of these two pairs. Increases in G were found mostly in dinucleotides such as GpA, that would contribute to an increase in purine sequences. The unchanged base composition found for G, in this RNA product, is due to decreases in G in two other dinucleotides.

The sixteen dinucleotides were divided into four groups of similar sequence , as follows: 1. Purine dinucleotides: ApA, ApG, GpA, GpG; 2. Pyrimidine dinucleotides: UpU, CpU, UpC, CpC; 3. Mixed, purine-pyrimidine, complementary pairs: CpA:UpG, ApC:GpU; 4. Mixed, purine-pyrimidine, self-complementary dinucleotides: UpA, ApU, CpG, GpC. The fraction found in each group, for the T7 RNA transcripts, is shown in Table 5E. Compared to the denatured T7 RNA transcripts, the native T7 RNA transcripts showed an increase in purine dinucleotides, a decrease in pyrimidine dinucleotides, an unchanged fraction of mixed, complementary dinucleotides. The ratio of purine/pyrimidine dinucleotides was 1.11 for the native T7 RNA transcripts and 0.98 for the denatured T7 RNA transcripts.

In contrast to the results found for T7 DNA transcription, analysis of the dinucleotide frequencies of the RNA transcripts of T4 DNA did not show a unique purine enrichment in the transcripts of native T4 DNA. The asymmetry ratios of the RNA transcripts of native and denatured T4 DNA were similar to each other (Table 6C, selected values shown below). Many of the complementary dinucleotide pairs, of both the native and denatured T4 RNA transcripts had asymmetry ratios significantly different from 1.0. In the transcripts of native DNA only one pair showed an increase in purine sequences (ratio: 1.26). One pair decreased (ratio: 0.93) and the other two had matching frequencies of purine and pyrimidine sequences. In

Selected values from Table 6C

Dinucleotide pair	Asymmetry Ratios (T4 DNA template)			
	Native	(p)	Denatured	(p)
ApA:UpU	0.93	0.001	1.08	0.001
GpA:UpC	1.10	0.10	1.20	0.001
ApG:CpU	0.88	0.10	1.00	0.50
GpG:CpC	1.26	0.001	1.44	0.001
CpA:UpG	0.72	0.001	0.73	0.001
GpU:ApC	1.17	0.10	1.22	0.001

the transcripts of denatured T4 DNA, three of these same four pairs showed an increase in purine sequences. The fourth pair had equal frequencies. Three of the mixed, complementary dinucleotide pairs had asymmetry ratios that were significantly different from 1.0. The variance of the asymmetry ratio, for both sets of transcripts, was much greater than the expected variance due to experimental error (Table 9), and therefore the asymmetry ratios of both sets of data were significantly different from 1.0. Transcription of both native and denatured T4 DNA, by these criteria, is therefore considered to be asymmetric.

The individual dinucleotide frequencies of the native and denatured T4 RNA transcripts were compared (Table 6D, selected values shown below). All the purine dinucleotides had lower frequencies in the RNA transcripts of the native DNA than in those of the denatured DNA. Many of the other dinucleotides had the same

Selected values from Table 6D

Dinucleotide Sequence	Frequency (RNA) T4 DNA template		Base change (in Native) + increase - decrease 0 no change
	Native	Denatured	
ApA	0.119	0.125	- A
GpA	0.0625	0.0673	- G,A
ApG	0.0506	0.0530	- A,G
GpG	0.0325	0.0344	- G
GpU	0.0498	0.0534	0
ApC	0.0426	0.0439	0
UpA	0.0858	0.0854	0
ApU	0.105	0.105	0

frequencies in both sets of RNA transcripts. When the dinucleotides were divided into four groups of similar sequence (Table 6E), the transcripts of native T4 DNA were found to have equal amounts of purine and pyrimidine dinucleotides. The transcripts of the denatured T4 DNA, however, had an increased purine, and a decreased pyrimidine fraction.

In contrast to the results found for T7 DNA transcription, the native T4 RNA transcripts are not enriched in purine sequences, and the denatured T4 RNA transcripts are not symmetric. The observation that transcription of denatured T4 DNA appears to be asymmetric, is consistent with the work of Brody and Geiduschek (1970), who have shown that *in vitro* transcription of denatured T4 DNA retains some of the specificity of transcription found for *in vitro* transcription of the native template. They also showed that the RNA transcribed from denatured T4 DNA was different from the RNA transcribed from native T4 DNA. The transcripts of denatured T4 DNA had a large amount of self-complementary RNA (i.e. they were more symmetric), and they also had different hybridization-competition properties. The difference between the results found for T4 and T7 DNA transcription could be due, in part, to the different length of the DNA template transcribed *in vitro*, in each phage system. The native T4 DNA template transcribed *in vitro* has a length equivalent to approximately 22×10^6 daltons, (Brody *et al*., 1970). The T7 DNA template transcribed *in vitro* is much smaller, being equivalent to approximately 2×10^6 daltons (Summers and Siegel, 1970). The ability to detect an enrichment of purine dinucleotides in the RNA transcripts depends, at least partly, on the relative length of the purine-rich regions. A short unique purine-rich region, for example, might be detected if it were 25% of the length of the total RNA, but 'diluted' out if it were only 5% of a much longer RNA transcript. Although no assumptions are being made here about the relative lengths of presumed pyrimidine-rich control regions of transcribing DNA strands, it is possible that the much longer T4 RNA transcripts might lead to a 'dilution' effect on purine-rich regions in the RNA. In order to test this possibility, another experiment was carried out with native T4 DNA, in which RNA synthesis was limited to a small region of the DNA. The limited-synthesis reaction was incubated at a lower temperature (23° C) and for a shorter time (6 minutes) than the standard enzyme reaction (Materials and Methods II). RNA synthesized under these conditions is the size of immediate early T4 message (approximately 4300 nucleotides long), as described by Witmer (1971).

Analysis of these limited-synthesis RNA transcripts showed that they were more asymmetric than the RNA transcripts of both native and denatured T4 DNA, discussed above, (Tables 6C, 7C, selected values shown below). Although the base composition and base ratios of this RNA did not differ noticeably from the two other T4 RNA transcripts (Tables 6B, 7B, selected values shown below), there was a significant enrichment of purine sequences in the limited synthesis RNA transcripts.

Selected values from Tables 6C, 7C, 6B and 7B

Dinucleotide pair			Asymmetry Ratios T4 DNA template			
	Native limited	p	Denatured	p	Native standard	p
ApA:UpU	1.05	0.20	1.08	0.001	0.93	0.001
GpA:UpC	1.43	0.001	1.20	0.001	1.10	0.010
ApG:CpU	1.12	0.020	1.00	0.50	0.88	0.10
GpG:CpC	1.73	0.001	1.44	0.001	1.26	0.001
CpA:UpC	0.52	0.001	0.73	0.001	0.72	0.001
GpU:ApC	1.48	0.02	1.22	0.001	1.17	0.10

DNA	Base Composition/Ratio				
	A	G	C	U	G/C
Native-limited	0.321	0.185	0.167	0.340	1.11
Denatured	0.328	0.188	0.157	0.328	1.20
Native	0.317	0.181	0.161	0.340	1.13

Compared to the standard native T4 RNA transcripts, there was a large increase in the asymmetry ratio for five of the complementary dinucleotide pairs. Purine sequences GpG and GpA were particularly enriched in the limited-synthesis RNA product. The variance of the asymmetry ratios of this RNA (24×10^{-2}) was much higher than the variance of the denatured DNA transcripts (7.9×10^{-2}) (Table 9). These transcripts were, therefore, highly asymmetric and enriched in purine sequences.

The individual dinucleotide frequences of the limited-synthesis transcripts were compared with the two other T4 RNA transcripts (Tables 6D, 7D, selected values shown below).

Selected values from Tables 6D, 7D

Dinucleotide sequence	Nearest-neighbor Frequency (RNA) T4 DNA template		
	Native (limited)	Denatured	Native (standard)
ApA	0.128	0.125	0.119
GpA	0.0731	0.0673	0.0625
UpC	0.0511	0.0558	0.0567
ApG	0.0588	0.0530	0.0506
GpG	0.0340	0.0344	0.0325
CpC	0.0197	0.0239	0.0258

Increases were found in the limited-synthesis RNA product for purine dinucleotides ApA, GpA and ApG. A decrease in pyrimidine dinucleotide CpC was also observed. The asymmetry ratio of the pair GpG:CpC, therefore increased, and the purine dinucleotide GpG was 'relatively' enriched in these transcripts. When the dinucleotides of the limited-synthesis RNA transcripts were divided into four groups of similar sequences, the purine dinucleotide fraction was high, and the pyrimidine dinucleotide fraction was low (Table 7E). The dinucleotide distribution was similar to the one found for the transcripts of denatured T4 DNA, except that the purine dinucleotide fraction was higher in the limited-synthesis RNA product. It is concluded, from the analysis of the RNA transcripts of T4 DNA, that: 1. In the standard synthesis reaction, both native and denatured T4 DNA templates are transcribed asymmetrically. The extent of the asymmetry is about the same in both types of transcription; 2. The T4 native DNA template, transcribed in the limited-synthesis transcripts, has a higher proportion of pyrimidine dinucleotides than the native DNA transcribed in the standard raction. Restricting transcription to a smaller region of the T4 DNA template has made it possible to detect pyrimidine-rich regions in the template that were not observable before; 3. Compared to the denatured T4 RNA transcripts, the higher asymmetry ratios and the higher variance of the limited-synthesis RNA transcripts, suggests that unique regions of the selectively transcribed DNA template are pyrimidine-rich.

The dinucleotide frequencies of T4 DNA have previously been determined by Josse et al. (1961). These values have been compared with the dinucleotide frequencies of the RNA synthesized by transcription of T4 DNA, determined in the present studies (Table 7D). The nearest-neighbor frequencies of the RNA purine dinucleotides were found, in most instances, to be higher than their complementary dinucleotides in the DNA. For example, ApA and GpA had higher frequencies in all the RNA transcripts than TpT and TpC in the DNA. Most of the pyrimidine dinucleotide frequencies were considerably lower in the RNA than in their DNA complements. The distribution of purine and pyrimidine dinucleotides in the T4 DNA and in the RNA transcripts of T4 DNA was different (Tables 6E, 7E). In the DNA the ratio of purine to pyrimidine dinucleotides was equal to 1.0, while in the RNA transcripts the values were as follows: native template: 0.99, denatured template: 1.09, limited-synthesis, native template: 1.20. This comparison of the dinucleotide frequencies of T4 DNA with the frequencies found for the RNA transcripts of T4 DNA confirms the earlier conclusion that pyrimidine sequences are preferentially transcribed in T4 RNA synthesis.

The results of analysis of the dinucleotide frequencies of transcripts of T5 DNA were similar to those found with T4 DNA (excluding the limited-synthesis product), in that the RNA products of both native and denatured DNA transcripts were weakly asymmetric (Table 8C, selected values shown below).

<u>Selected values from Table 8C</u>

Dinucleotide pair	Asymmetry Ratios			
	Native	(T5 DNA template)	Denatured	
		p		p
ApA:UpU	1.02	0.20	0.90	0.10
GpA:UpC	1.11	0.001	0.92	0.005
ApG:CpU	0.99	0.40	1.07	0.05
GpG:CpC	1.15	0.001	1.22	0.001
CpA:UpG	0.77	0.001	0.69	0.001
GpU:ApC	1.10	0.001	1.11	0.001

In the RNA product of denatured T5 DNA transcription, the first four matching dinucleotide pairs included two pairs with asymmetry ratios greater than 1.0, one pair with a ratio less than 1.0, and one pair with a ratio equal to 1.0. In the native T5 RNA transcripts, two pairs had ratios greater than 1.0 and two had ratios equal to 1.0. There was a relative enrichment of purines,predominantly G, in both sets of transcripts, but the asymmetry was less marked in these transcripts than it had been in the T4 and T7 RNA transcripts. The two matching pairs of mixed complementary dinucleotides had asymmetry ratios different from 1.0, in both sets of T5 RNA transcripts. The observed variance of both sets of asymmetry ratios, for the T5 transcripts, was found to be lower than the values found for any of the asymmetric transcripts. The asymmetry ratios, however, were found to be different from 1.0, and both types of transcripts, by this criterion, were considered to be asymmetric.

The individual dinucleotide frequencies of both T5 RNA transcripts were compared (Table 8D, selected values shown below).

Selected values from Table 8D

Dinucleotide sequence	Frequency (RNA) T5 DNA template Native	 Denatured	Base change (in Native) + increase - decrease
ApA	0.102	0.0950	+ A
GpA	0.0606	0.0524	+ G,A
ApG	0.0632	0.0680	- A,G
GpG	0.0426	0.0455	- G
UpG	0.0644	0.0725	- G,U
ApU	0.0907	0.0864	+ A,U

Two purine dinucleotides had higher frequencies in the native T5 transcripts, and the two others had lower frequencies. The pyrimidine dinucleotides of both transcripts were similar to each other, but almost all the mixed dinucleotides were different from one another.

Comparing the distribution of RNA in purine, pyrimidine and mixed dinucleotides (Table 8E), the purine and pyrimidine dinucleotides were found to be equal to each other in the transcripts of denatured DNA. In the transcripts of native T5 DNA, the purine fraction was a little higher than the pyrimidine fraction. T5 DNA transcription appears to resemble T4 transcription. Based on the asymmetry ratios, there is some specificity retained in the *in vitro* transcription of the denatured template, as in T4 DNA transcription. There is also a limited enrichment of purines in the transcripts of native T5 DNA.

The nearest-neighbor frequencies of T5 phage DNA have been determined (Swartz *et al*. 1962), and the results found in these studies, for the nearest-neighbor frequencies of RNA transcripts of T5 DNA, were compared with them (Table 8D). Many of the RNA dinucleotide frequencies were different from those in the T5 DNA. However, there was no consistent pattern in the differences found. Only one purine dinucleotide, GpG, had a higher frequency in both types of RNA transcript than its DNA complement CpC. This is consistent with the high G/C values found for the T5 RNA (Table 8B). The pyrimidine dinucleotide CpU was also higher in the RNA transcripts, than its DNA complement ApG. When the dinucleotides were divided into four groups of similar sequence, the distribution in the T5 DNA of the purine, pyrimidine and mixed dinucleotides was different than that found for the RNA products (Table 8E). The ratio of purine to pyrimidine dinucleotides, which was 0.96 in the DNA, was 1.05 for the transcripts of native DNA and 0.99 for the transcripts of denatured T5 DNA. The mixed, self-complementary dinucleotides were lower in the RNA than in the T5 DNA. Transcription of both native and denatured T5 DNA is selective to some extent, since the sequence pattern of the transcripts is different from that of the T5 DNA. There is only a suggestion in these results, however, that enriched pyrimidine sequences in the DNA are transcribed.

The dinucleotide frequencies of RNA transcribed from f1 RF DNA template were also determined, as part of the experiment to establish the selectivity of the RNA polymerase used in these transcripts (Results III). Phage f1 was not studied in the same way as the T phages, but the partial results are presented in Table 4C. These transcripts were found to be extremely asymmetric, the direction of the asymmetry being reversed, with low asymmetry ratios, such as 0.57 and 0.60. The DNA template here is very rich in purine sequences, and this suggests that perhaps a different kind of transcriptional control may be involved.

Non-random nature of nearest-neighbor frequencies of the RNA transcripts. The nearest-neighbor frequencies of the RNA transcripts, determined in these studies, have been analyzed for deviation from the nearest-neighbor frequencies predicted for random arrangement of the mononucleotides in the RNA. In a randomly arranged RNA, the frequency of any nearest-neighbor pair should be predictable as the product of the frequencies of its constituent mononucleotides; for example, the frequency of UpA would be the same as the frequency of ApU, and would be equal to the product of the frequencies (or mole %) of Up and Ap. Using the known base composition of each DNA studied, the random nearest-neighbor frequencies of all the dinucleotides were calculated in this way, for the RNA transcripts of T7, T4 and T5 DNA (Tables 5D, 7D and 8D, respectively). The deviation of the dinucleotide frequencies from their random values was analyzed in three ways, as follows: 1. the sixteen dinucleotide frequencies of each RNA synthesized were tested against their random values; 2. Each dinucleotide sequence was tested against its random value for all the RNA transcripts synthesized; 3. The frequency of each individual dinucleotide sequence was compared to its random value.

The variance (Appendix B) of the observed dinucleotide frequencies from the predicted (random) values was calculated for the sixteen dinucleotide sequences of each RNA transcript (Table 10). The expected variance of the predicted random frequencies was determined from the experimental error of each experiment. The ratio of the two variances, and the number of degrees of freedom, determine the level of probability, P, at which the observed frequencies are significantly different from the random frequencies. For the purposes of this study, if P is greater than 0.05 the observed and random frequencies are taken to be the same. If P is less than or equal to 0.05, the observed frequencies are considered to be different from the random ones. The value of P was found to be less than 0.05 for each RNA transcript, therefore the nearest-neighbor frequencies were determined to be non-random for all the RNA transcripts.

In order to detect any similarity in the deviation from random values of particular dinucleotides, the deviation of each dinucleotide was determined for all the RNA transcripts (Table 11). The variance (Appendix B) of the observed frequencies from the predicted values was calculated for each dinucleotide. The variance expected in the predicted value, due to experimental error, was also calculated. The level of probability, P, at which the observed frequencies are significantly different from the random values was determined, as described above. All the sets of individual dinucleotide frequencies, except GpU, were found to be different from the predicted random values. The random finding for GpU is due to some unusually scattered data values, which resulted in a relatively high value for the variance due to experimental error. Dinucleotides with lower observed variances, such as 0.27×10^{-4} for ApG, do not differ as much from random values as others, such as ApA, for which the variance was 3.2×10^{-4}. When deviation of the individual dinucleotides was compared to this average calculation, it was found (Table 11) that 41% of the dinucleotides had values considered to be equal to the random values, and about 30% were above or below random values.

The asymmetry previously described in analysis of the RNA transcripts is reflected here in the way in which many of the matching dinucleotide pairs varied from the random values. For example, in the pair CpA/UpG the values of CpA were equal to or less than the random values, while for UpG all the values were above the random values. In the four purine dinucleotides, almost all the frequencies were equal to, or greater than the random values, while in pyrimidine dinucleotides UpC and CpC the frequencies were equal to, or less than random values. The non-random nature of the dinucleotide frequencies of the RNA transcripts can also be seen in a comparison of the frequencies of isomeric pairs, such as ApG-GpA and UpC-CpU. The calculated random value of each member of the pair is the same, since it is determined from the product of the mole % of each base in the dinucleotide. The dinucleotide frequencies found experimentally for many of these isomeric pairs, however, were quite different from each other. For example, in the denatured T4 RNA transcripts (Table 6C), the frequency of ApG was 0.0530 while GpA was 0.0673. Similar differences were found in all the RNA transcripts studied. The dinucleotide frequencies of several bacterial and viral DNAs have also been shown to be non-random (Josse et al., 1961; Swartz et al., 1962).

The way in which the observed dinucleotide frequencies varied from the predicted random values has been presented graphically, according to the method of Kaiser and Baldwin (1962) in Figures 4-7. On each graph the observed nearest-neighbor frequency of a specific dinucleotide pair, for each DNA studied, was plotted against the value predicted from a random distribution. If the observed and predicted values were equal, then the points would fall on the straight 45 degree line, drawn on each graph. The points plotted, therefore, fall above, on, or below this line, as the observed dinucleotide frequency is greater than, equal to, or less than the predicted random value. For phages T4 and T5 the nearest-neighbor frequencies of the DNA complements have also been plotted on each graph.

For some of the dinucleotides the frequencies can be seen to be very different from each other, corresponding to the AT/GC values of the DNA of each phage (AT/GC = 1.86 for phage T4, 1.53 for phage T5 and 1.08 for phage T7; Wyatt and Cohen, 1953; Lunan and Sinsheimer, 1956). For example, on the graphs of dinucleotides ApA and UpU, the points are widely scattered. For other dinucleotides, such as GpA and UpC, the frequencies were similar and the points are clustered on the graphs. The DNA values generally fell closer to the random 45 degree line than the RNA values.

The dinucleotide frequencies of the pair ApA:UpU were considerably higher than the random values, for the RNA transcripts of native DNA, and for the DNA template as well. In phages T4 and T5, both high in AT base composition, this pair, ApA:UpU, has the highest predicted random frequency, and yet there was a large additional increase over the random values, in both the DNA and RNA frequencies. The dinucleotide pair GpG:CpC was quite different in both DNA and RNA dinucleotide frequencies. The DNA frequencies were close to the random values for both dinucleotides. In the RNA, GpG remained at random levels, but CpC was found at lower than random frequencies. These observations are in agreement with studies of the distribution of thymine and cytosine oligonucleotides in bacterial and viral DNA (Mushynski and Spencer 1970a,b; Rudner *et al*., 1972). In *B*. *subtilis* DNA the distribution of thymine oligonucleotides was found to be greater than expected from a random distribution, while cytosine oligonucleotides followed the expected random distribution (Rudner *et al*., 1972). In bacterial and viral DNAs, long cytosine oligonucleotides, greater than 5 in chain length, are quite rare, whereas long thymine oligonucleotides are frequently found. In these studies, dinucleotide ApA, which is complementary to thymine sequences, was found at higher than random frequencies. Dinucleotide GpG, complementary to cytosine sequences, was found to be randomly distributed.

The frequencies of the purine:pyrimidine dinucleotide pairs, ApG: CpU and GpA:UpC, were somewhat similar to each other (Figure 5a-d). Many of the dinucleotides in this group were at, or near, random frequencies. In the RNA transcribed from native DNA, purine dinucleotide

Discussion

In these studies, the proposal that pyrimidine-rich regions of transcribing DNA strands are involved in transcriptional control has been examined. These pyrimidine-rich regions would be 'recognition' sites for specific interaction of the RNA polymerase and the DNA template. The specificity of the interaction would be determined by the DNA sequence, and could signal any step in the transcription process. Such a control sequence would not necessarily have to be transcribed into an RNA product. However, a sequence involved in initiation control, for example, could be close to or within the early region transcribed. It was proposed, therefore, in the present studies, to determine if transcription of such pyrimidine-rich regions could be detected in the RNA transcripts. Limited RNA synthesis, confining transcription to very early mRNA, was also used, in one instance.

In order to see if these pyrimidine-rich regions are transcribed into the RNA product, the dinucleotide frequencies of RNA synthesized _in vitro_ on T4, T5 and T7 phage DNA templates were determined. It was found that the RNA is enriched in purine dinucleotides and therefore that pyrimidine-rich regions are preferentially transcribed. The increased proportion of purine dinucleotides in the RNA product was most pronounced in the RNA transcripts of T7 DNA and in the limited-synthesis transcripts of T4 DNA. In the RNA transcripts of T5 DNA, and for standard synthesis of T4 DNA, there was a limited enrichment of purine sequences.

found (Hinkle et al., 1972).

It was assumed that the RNA transcripts of denatured DNA, in these studies, would be quite symmetric in composition and matching dinucleotide sequences. Some of them, however, were found to be asymmetric and similar to the transcripts of native DNA (Tables 6C -8C). It had to be concluded, from the extent of asymmetry found, that transcription of denatured T4 and T5 DNAs had retained a significent amount of specificity. The nearest-neighbor frequencies of the T7 RNA transcripts were consistent with the initial assumption, of symmetric transcription, since the asymmetry ratios of the complementary dinucleotide pairs were found to be equal to 1.0 (Table 5C). The asymmetry found in the RNA transcripts of denatured T4 and T5 DNA made it more difficult to evaluate the extent of enrichment of purine dinucleotides in their respective transcripts of native DNA. In the case of T4 transcription, this difficulty was overcome in the limited-synthesis experiment, where a definite enrichment of purine sequences was found in the shorter transcripts of native T4 DNA (Table 7C). As was discussed earlier, in the case of the T4 DNA, it is possible that the results found for the native T5 DNA may be due to the size of the in vitro transcripts made. T5 DNA is somewhat smaller than T4 DNA (83 x 10^6 daltons, Rubenstein, 1968), and in vitro transcription of native T5 involves a total length of DNA estimated to be about 20 x 10^6 daltons (Pispa and Buchanan, 1971). (The T4 RNA transcript was estimated to be about 22 x 10^6 daltons). The size of the RNA transcripts may make it difficult to detect small unique pyrimidine-rich regions by this kind of analysis.

On the other hand, pyrimidine sequences may not be involved in the control of transcription of T5 DNA. The T5 DNA template differs from both T4 and T7, for example, in that it contains six or more genetically distinct single-stranded breaks, on one of the DNA strands (Bujard and Hendrickson, 1973). These breaks could be involved in transcriptional control and therefore require different DNA sequences for enzyme recognition.

The dinucleotide frequencies of the DNA of phages T4 and T5 have been determined (Josse _et al_., 1961; Swartz _et al_., 1962) and it was therefore possible to compare the dinucleotide frequencies of their RNA transcripts with the values obtained for the DNA template (Tables 6D, 7D, 8D). When compared to the dinucleotide frequencies of the DNA template a preferential transcription of purine dinucleotides was found for all the T4 RNA transcripts. In the native and denatured T5 RNA transcripts there was an increase in dinucleotide GpG.

The present studies show the selective transcription of pyrimidine sequences on the transcribing DNA strand but do not provide any basis of distinguishing between short runs of pyrimidines (chain length 5-6) and clusters (chain length 8-13) in the DNA template. Cytosine clusters were proposed earlier as control elements in transcription, based on the binding of transcribing DNA strands to poly (U,G) (Szybalski _et al_., 1966). When pyrimidine isostichs were isolated from viral and bacterial DNA, long cytosine clusters were not found, but long thymine and mixed cytosine-thymine oligonucleotides (chain length 13) were found (Mushynski and Spencer, 1970a,b; Rudner _et al_., 1972; Rudner and LeDoux, 1974). These results strengthened the earlier proposal that pyrimidine clusters might be involved in initiation of RNA synthesis. Studies of oligoribonucleotidebinding to homologous phage DNAs suggest that specific binding of poly G and guanine-rich ribopolymers requires oligomers of at least 10 residues in chain length (Niyogi, 1969; Niyogi, 1973). When cytosine clusters of this length were not found some alternative proposals for binding were made. It was suggested that the binding could be due to C-rich DNA that consists of small C clusters, separated by a single purine residue (Mushynski and Spencer, 1970b). Poly G can also form complexes with segments containing short TC sections (Morgan and Wells, 1968).

Several initiation and termination regions in E. coli , and in several viruses, have been sequenced (Gilbert and Maxam, 1973; Sekiya and Khorana, 1974; Maniatis et al., 1974; Pribnow, 1975; Schaller et al., 1975). These regions are representative of other control regions for E. coli RNA polymerase, since all E. coli, and many viral genes are believed to be transcribed by one RNA polymerase enzyme which may be modified under different conditions (Fukuda et al., 1974; Wickner and Kornberg, 1974). The sequences involved in termination and initiation would be expected, therefore, to have some similarities for recognition by the E. coli enzyme. Two common sequences of six bases, recognized by the Hind endonuclease,(GTCGAC and GTTAAC), have been found in promoter sites for E. coli RNA polymerase, in DNA of bacteriophages T7 and lambda, and in simian virus 40 and human adenovirus (Allet et al., 1974). In the polymerase protected region of the T7 A_3 promoter, a 7 base sequence has been found (5'-TATPuATG-3': Pribnow, 1975) which is homologous with promoter sequences in simian virus 40 (Dhar et al., 1974), lambda P_L promoter (Maniatis et al., 1974), E. coli Tyr-tRNA gene (Sekiya and Khorana, 1974), the lac region (Dickson et al., 1975) and fd promoter-I (Schaller et al., 1975). In the control regions already sequenced, no long pyrimidine isostichs (chain length 12-13) have been found. Although such isostichs are known to be present in some of these DNAs, e.g. in T7, lambda and E. coli DNAs (Mushynski and Spencer, 1970a,b; Rudner and LeDoux, 1974), they do not appear to be present in the initiation and termination regions.

The secondary structure of the DNA helix depends on its base composition (Bram, 1971; Bram and Tougard, 1972). For example, natural DNAs that are very rich in AT (AT/GC greater than 2.0) have a different secondary structure than DNA of moderate and low AT content (Bram, 1971). It is also quite possible for the secondary structure of the DNA to provide a basis for specific interactions, such as RNA polymerase binding to a promoter site, if the secondary structure exists in a conformation which is dependent on the primary sequence of the DNA (Bram 1973). Short regions of highly regular base sequence can also modify the secondary structure of the DNA helix in a limited region, and provide a recognition site for transcriptional control. In diffraction studies on the conformation of DNA (Langridge, 1969) it was found that DNA sequences which contain only pyrimidines on one strand and purines on the other, have X-ray diffraction patterns that all differ from native DNA. Poly dA;dT has all the features of a helical structure, but with a different base tilt; poly d(T-C); poly d(G-A) also differs clearly from the normal DNA pattern and poly dG:poly dC had the most unusual structure, appearing to be non-integral. It was concluded (Langridge, 1969) that DNAs, containing local regions of highly regular sequence, can clearly adopt unusual configurations. Such a modified DNA structure could then serve as a signal for transcriptional control. The enrichment of pyrimidine sequences in the transcribing DNA strands, found in the present studies, may involve just such short regular regions. These regular regions may cause local modification of the helical structure of the DNA, providing the configurational changes necessary for recognition signals.

DNA sequences, which have been determined for a number of control regions (Pribnow, 1975; Schaller *et al.*, 1975) were analyzed for their pyrimidine enrichment and for short regions of regular base sequence. The frequencies of purine, pyrimidine, mixed complementary and mixed self-complementary dinucleotides, were determined for each region, and compared with the dinucleotide frequencies reported here. The control sequences analyzed include *E. coli lac* operator-promoter region (Gilbert and Maxam, 1973), the promoter and terminator regions of *E. coli* Tyr-tRNA gene (Sekiya and Khorana, 1974; Loewen *et al.*, 1974), the lambda P_L promoter (Maniatis *et al.*, 1974), the T7 A_3 promoter (Pribnow, 1975), simian virus 40 (Dhar *et al.*, 1974) and fd region-I promoter (Schaller *et al.*, 1975) (Table 12). The *lac* operator-promoter region (Gilbert and Maxam, 1973), which is known to be transcribed, was found to have a higher frequency of pyrimidine dinucleotides (0.35 for the operator region, 0.32 for the operator-promoter region) than *E. coli* DNA (0.243; Josse *et al.*, 1961). Both *lac* repressor and RNA polymerase bind to different regions of this DNA. The frequency of purine dinucleotides is very low (0.10 and 0.12) compared to the *E. coli* value (0.237). There are two sequences in the region that are pyrimidine-rich. There is one sequence of

13 bases, consisting of 10 pyrimidines, in clusters of 2 and 3, (CTCGCCTATTGTT), and one sequence of 9 bases, consisting of 8 pyrimidines (TCCTTTGTC).

In the *E. coli* Tyr-tRNA gene, promoter and terminator regions (Sekiya and Khorana, 1974; Loewen *et al.*, 1974) are probably not transcribed. The dinucleotide frequencies were determined, in this analysis, for the strand which is continuous with the transcribed tRNA gene. In the promoter region, both the pyrimidine and purine dinucleotides have very low frequencies (0.11 and 0.21, respectively). The mixed self-complementary dinucleotides have an unusually high value (0.50). The terminator region of the gene has a high frequency of purine dinucleotides (0.45). The combined frequency of purine and pyrimidine dinucleotides in this region is 0.72, suggesting that there is a regular local structure. The terminator region consists of alternating purine and pyrimidine clusters, 2 to 6 bases long (AGTGAAAGTTTTCAGGGGACTTGA). RNA synthesized from this region would be pyrimidine-rich. In several phage T7 RNAs, a long pyrimidine region, containing a sequence of six pyrimidine residues, (CCCUUUAU-OH), was found in the terminal sequences (Kramer *et al.*, 1974) of four of five mRNA species studied. This suggests that small purine-rich DNA regions may be involved as termination signals for transcription.

The P_L promoter binding site in lambda, which also binds repressor (Maurer *et al.*, 1974), was found to have an unusually high frequency of complementary and self-complementary mixed dinucleotides (0.38, 0.31), compared to the *E. coli* values (0.25, 0.27) (Table 12). This region is not transcribed *in vivo* (Maniatis *et al.*, 1974). There is one pyrimidine-rich sequence, 8 residues long, containing 6 pyrimidines (TCACCGCC) in this control region. In the *E. coli* RNA polymerase binding and initiation site in phage fd (Sugimoto *et al.*, 1975) the sum of purine and pyrimidine dinucleotide frequencies is high in the binding site, and in the shorter transcribed region, as well (0.60 and 0.67, respectively). These frequencies have been compared to the frequencies determined in this study, for f1 RF transcripts, since values for the fd single-stranded DNA are not available. The distribution pattern of the purine, pyrimidine and mixed dinucleotide frequencies in this control region is different from the pattern of the RNA transcripts. The site of initiation of transcription is in the center of a pyrimidine-rich region, 17 residues long, (TTATCTGTCC$\overset{*}{\text{C}}$ATTTCT); transcription is to the right, starting at the asterisk. There is also a purine-rich region in the early part of the transcript, (GGACTAAAAACTAAA).

In the phage T7 A_3 promoter region (Pribnow, 1975), there is an increase in pyrimidine dinucleotides (0.30; T7 value: 0.257) and a decrease in purine dinucleotides (0.05; T7 value: 0.253) (Table 12). There are two pyrimidine-rich sequences in the region. One is at the beginning of transcription, and consists of a sequence of 14 residues, containing 11 pyrimidines, (CTTTGCTGTCACTC). The other is at the left end of the polymerase-protected region, and consists of 6 residues, (TTCATTT). In the simian virus 40 promoter region (Dhar *et al.*, 1974), there is one pyrimidine-rich sequence which is transcribed, and is at the initiation site for transcription. It contains 11 residues, of which 9 are pyrimidines, (TTTATTTCGTT).

In these regulatory sequences, the dinucleotide frequencies are quite different from the frequencies of the DNA genomes. DNA base composition in control regions is also very different from the base composition of the genome. For example, the E. coli RNA polymerase binding sites in T5, T7 and lambda are alternating high GC (about 14 base pairs long) and high AT regions (about 38 nucleotides long) (Giacomoni et al., 1974). In all of the sequences analyzed above, there are short regions of highly regular sequences which may be able to modify the local secondary structure of the DNA. Such configurations could be involved in determining specificity in transcriptional control.

The repressor binding capacity of the lactose operator in E. coli has been studied, to determine if natural DNA has the same conformation throughout its entire length, or if different structures exist at various sites along the chain (Chan and Wells, 1974). The ability of single-strand specific nucleases, and of non-specific nicking agents, to abolish repressor binding was determined. Certain regions of the lactose operator were quite sensitive to single-strand specific nucleases and therefore were assumed to have different structures than the greater part of the DNA. The nature of the uniqueness is not known, but several models were proposed: 1. The region has a low thermostability and could be in an open structure for a large part of the time; 2. The region has 2-fold symmetry and could form a cruciform structure, each strand looping out on itself; 3. The DNA may have a different configuration than normal helical DNA, such as a different base tilt or altered helical dimensions.

In conclusion, the results of the present study are consistent with a model of transcriptional control which proposes that conformational changes in the secondary structure of the DNA helix, based on the primary sequence in the DNA, allow for the formation of unique structures which could participate in highly specific interactions with proteins, and other regulatory molecules. The possibilities for specific modulation of the secondary structure by base composition, in such a system, would be very extensive. Pyrimidine-rich regions, of varying size and cluster lengths, and alternating high GC and AT-rich regions, would all affect the secondary structure in their own unique ways. This would provide the basis for a large number of different, highly specific interactions with proteins and other regulatory molecules. These conformational changes would also be very sensitive to environmental conditions, such as ionic strength and relative humidity (Bram and Tougard, 1972). The findings in the present study suggest that purine and pyrimidine sequences, of 2 to 7 residues in length, occurring in a regular pattern in control regions of the DNA, provide one basis for unique local structural modification, and for control of RNA synthesis.

Appendix A: Calculation of nearest-neighbor frequencies, according to Josse et al. (1961)

Following the method outlined in Materials and Methods III, ^{32}P measurements of the four 2'-(3')-nucleotides were made. Samples of this data, for each phage studied, are given in Tables 4A, 5A, 6A, 7A and 8A. For each electrophoretic separation the fraction of the total ^{32}P label that was transferred from the 5'-nucleotide to the 2'-(3')-nucleotides was determined. The frequency of each 3'-nucleotide present in the RNA product depends on the relative frequency with which the particular 5'-nucleotide, from which the label is transferred, is incorporated into the RNA. If the base composition of the RNA is known then each fractional transfer value is multiplied by the base composition of the particular 5'-nucleotide to give the nearest-neighbor frequency of the 2'-(3')-nucleotide. If, for example, in a reaction using ATP^{32} as substrate, the fraction of total label transferred from pA to Gp is found to be 0.263, and if the base composition of the RNA is 0.279 Mole % for A, then the nearest-neighbor frequency of Gp or of dinucleotide GpA is equal to 0.263 x 0.279 or 0.0734. The base composition used for these calculations has been derived from the radioactivity measurements in the experiment and is the base composition of the RNA product made.

For the labeled substrate used in any one reaction the total 5'-nucleotide incorporated is equal to the total amount recovered as 2'-(3')-mononucleotides, i.e.,

$$\underset{\text{incorporated}}{\text{Total A}} = \text{ApA} + \text{ApU} + \text{ApC} + \text{ApG}$$

or $\Sigma A = \alpha A + \delta U + \gamma C + \lambda G$

where α, δ, γ and λ are the fractions transferred from each substrate to Ap. Using the results obtained in the native T7 DNA reaction, Table 5C, we then have the following simultaneous equations:

$$
\begin{aligned}
1.00\,A &= 0.288\,A + 0.256\,G + 0.242\,C + 0.227\,U \\
1.00\,G &= 0.263\,A + 0.234\,G + 0.283\,C + 0.239\,U \\
1.00\,C &= 0.244\,A + 0.224\,G + 0.172\,C + 0.297\,U \\
1.00\,U &= 0.206\,A + 0.206\,G + 0.242\,C + 0.227\,U
\end{aligned}
$$

Using determinants, we may solve for any three of the unknowns in terms of the fourth. For this set of equations the solution is:

$A = 1.13\,U, \quad G = 0.946\,U, \quad C = 0.977\,U$

These values represent the proportion of each base present in the RNA and the sum of the four bases may be set equal to 1.0, i.e.,

$A + G + C + U = 1.00$

and therefore, $1.13\,U + 0.946\,U + 0.977\,U + 1.00\,U = 1.00$. Solving this equation for U we get the base composition of the RNA product made. From this equation, for T7 DNA, the solution is $U = 0.247$, $A = 0.279$, $G = 0.234$, $C = 0.241$. Multiplying the values determined, for the fractions transferred of each dinucleotide, by these values, the 16 nearest-neighbor frequencies may then be calculated. An example for native T7 DNA is given in Table A.

Table A. Nearest-neighbor base frequencies of native T7 DNA transcripts: 2'-(3')-nucleotide fraction x Mole % of 5'-nucleotide substrate

^{32}P-XTP	2'-(3') XMP	Sequence	Mole % ^{32}P-XTP	Fraction transferred	Nearest-neighbor frequency
ATP	AMP	ApA	0.279	0.288	0.0804
	GMP	GpA	"	0.263	0.0734
	CMP	CpA	"	0.244	0.0681
	UMP	UpA	"	0.205	0.0575
GTP	AMP	ApG	0.234	0.256	0.0599
	GMP	GpG	"	0.234	0.0545
	CMP	CpG	"	0.224	0.0524
	UMP	UpG	"	0.286	0.0669
CTP	AMP	ApC	0.241	0.302	0.0728
	GMP	GpC	"	0.283	0.0682
	CMP	CpC	"	0.172	0.0414
	UMP	UpC	"	0.242	0.0583
UTP	AMP	ApU	0.247	0.238	0.0588
	GMP	GpU	"	0.238	0.0590
	CMP	CpU	"	0.297	0.0734
	UMP	UpU	"	0.226	0.0684

Appendix B: Calculation of Variance

Variance is a measure of the variation found in a set of values. In any set of observed values there is a variance due to experimental error which represents the minimum variation of the data. The variation of a set of values from a predicted set of values may also be measured by variance. A comparison of these two variances may be used to determine whether or not the observed values are significantly different from the predicted ones. The level of probability of significant difference is determined from the ratio of the two variances and the number of degrees of freedom involved. (Table for F test for equality of variances: Snedocor).

The variance of a set of observed values, due to experimental error may be expressed as:

$$\text{Variance} = \frac{\Sigma D^2}{n-1}$$

where D = the difference between the observed value and the mean, and n = number of entries. The variance of a predicted set of values can also be determined from the experimental error of observed values to which they are being compared. To determine this variance the coefficient of variation of the observed values is calculated:

Coefficient of variation (v) =

$$\frac{\text{standard deviation}}{\text{mean}} = \frac{(\Sigma D^2/n-1)^{1/2}}{\Sigma\text{observed values}/n} =$$

the standard deviation of the experiment expressed as a percentage.

To calculate the variance, then, we have:

$$\text{Variance}_1 = \frac{\Sigma (vf)^2}{n-1}$$

where f = the predicted value, v = the coefficient of variation.

The variance of any set of values, from a known or predicted set, may be expressed as:

$$\text{Variance}_2 = \frac{\Sigma d^2}{n-1}$$

where d = the difference between the observed value and the predicted value, n is the same as before (Swartz et al., 1962).

In Table 9 the variance of the asymmetry ratios from the predicted value of 1.0, for four and six pairs of matching dinucleotides, has been calculated for each native and denatured DNA template used. For each DNA the experimental variance of the predicted value of 1.0 has been determined using the coefficient of variation, calculated for each of the sets of matching dinucleotide pairs. The following formulas were used:

$$\text{Variance}_1 = \frac{\Sigma (v \times 1.0)^2}{n-1}$$

$$\text{Variance}_2 = \frac{\Sigma (\text{Asymmetry Ratio} - 1)^2}{n-1}$$

where v = the coefficient of variation, 1.0 = the predicted asymmetry ratio and n = number of entries. The level of probability of significant difference (P) between the asymmetry ratios and 1.0 has been determined by the F test for the ratio of $\text{Variance}_1/\text{Variance}_2$.

In Tables 10 and 11 the variance of experimentally determined nearest-neighbor frequencies from the predicted random values has been determined. In these Tables:

$$\text{Variance}_1 = \frac{\sum (\text{observed frequency} - \text{random frequency})^2}{n-1}$$

and

$$\text{Variance}_2 = \frac{\sum (v \times \text{random frequency})^2}{n-1}$$

where the observed frequency = the experimentally determined dinucleotide frequency and the random frequency = dinucleotide frequency calculated from the phage DNA base composition.

Bibliography

Allet, B., Robert, R. J., Gesteland, R. F. and Solem, R. 1974. Class of promoter sites for Escherichia coli DNA-dependent RNA polymerase. Nature 249:217.

Berg, D., Barrett, D. and Chamberlin, M. 1971. Purification of two forms of Escherichia coli RNA polymerase and of sigma component. Methods in Enzymol. Grossman, L. and Moldave, K. eds. Academic Press, New York 21:506.

Bishop, J. O. 1969. Transcription of denatured T4 DNA. Biochim. Biophys. Acta 174:636.

Borst, P. and Aaij, C. 1969. Identification of the heavy strand of rat-liver mitochondrial DNA as the messenger strand. Biochem. Biophys. Res. Commun. 34:358.

Bram, S. 1971. Secondary structure of DNA depends on base composition. Nature New Biology 232:174.

Bram, S. 1973. Recent DNA fiber diffraction studies and their implications for DNA recognition. Cold Spring Harbor Symp. Quant. Biol. 38:83.

Bram, S. and Tougard, P. 1972. Polymorphism of natural DNA. Nature New Biology 239:128.

Brody, E. N. and Geiduschek, E. P. 1970. Transcription of the bacteriophage T4 template. Detailed comparison of *in vitro* and *in vivo* transcripts. Biochemistry 9:1300.

Brody, E. N., Diggelmann, H. and Geiduschek, E. P. 1970. Transcription of bacteriophage T4 template. Obligate synthesis of T4 prereplicative RNA *in vitro*. Biochemistry 9:1289.

Bujard, J. and Hendrickson, H. E. 1973. Structure and function of the genome of coliphage $T5^+$: 1. The physical structure of the chromosone of $T5^+$. Eur. J. Biochem. 33:517.

Burgess, R. R. 1969. A new method for the large scale purification of *Escherichia coli* deoxyribonucleic acid-dependent ribonucleic acid polymerase. J. Biol. Chem. 244:6160.

Burgess, R. R., Travers, A. A., Dunn, J. J. and Bautz, E. K. F. 1969. Factor stimulating transcription by RNA polymerase. Nature 221:43.

Chamberlin, M. and Berg, P. 1964. Mechanism of RNA polymerase action: Formation of DNA-RNA hybrids with single-stranded templates. J. Mol. Biol. 8:297.

Chan, H. W. and Wells, R. D. 1974. Structural uniqueness of lactose operator. Nature 252:205.

Cohen, S. N. and Hurwitz, J. 1967. Transcription of complementary strands of phage lambda DNA *in vivo* and *in vitro*. Proc. Nat. Acad. Sci. USA 57:1759.

Dause, J., Sentenac, A. and Fromageot, P. 1972. Eur. J. Biochem. 26:43. Interaction of RNA polymerase from *Escherichia coli* with DNA: Selection of initiation site on T7 DNA.

Dawid, E. and Chase, J. W. 1972. Mitochonrdial RNA in *Xenopus laevis* II. Molecular weights and other physical properties of mitochondrial, ribosomal and 4s RNA. J. Mol. Biol. 63:217.

Dhar, R., Weissman, S. M., Zain, B. S. and Pan, J. 1974. Nucleic Acids Res. 1:595. The nucleotide sequence preceding an RNA polymerase initiation site on SV40 DNA. Part 2. The sequence of the early strand transcript.

Dickson, R., Abelson, J., Barner, W. and Reznikoff, W. 1975. Science 187:27. Genetic regulation. The *lac* control region.

Dunn, J. J. and Studier, F. W. 1973. T7 early RNAs are generated by site-specific cleavage. Proc. Nat. Acad. Sci. USA 70:1559.

Fukuda, R., Iwakura, Y. and Ishihama, A. 1974. Heterogeneity of RNA polymerase in *Escherichia coli*. I. A new holoenzyme containing a new sigma factor. J. Mol. Biol. 83:353.

Geiduschek, E. P. and Grau, O. 1970. In Lepetit Coll. Biol. Med. I. ed. by L. Sylvestri, p. 190. Amsterdam:north-Holland Publ. Co. T4 anti-messenger.

Geiduschek, E. P. and Haselkorn, R. 1969. Messenger RNA. Ann. Rev. Biochem. 38:647.

Geiduschek, E. P., Nakamoto, T. and Weiss, S. B. 1961. The enzymatic synthesis of RNA: complementary interaction with DNA. Proc. Nat. Acad. Sci. USA 47:1405.

Geiduschek, E. P., Tocchini-Valentini, G. P. and Sarnat, M. T. 1964. Asymmetric synthesis of RNA *in vitro*: dependence on DNA continuity and conformation. Proc. Nat. Acad. Sci. USA 52:486.

Giacomoni, P. U., Le Talaer, J. Y. and Le Pecq, J. B. 1974. *Escherichia coli* RNA polymerase binding sites on DNA are only 14 base pairs long and are located between sequences that are very rich in AT. Proc. Nat. Acad. Sci. USA 71:3091.

Gilbert, W. and Maxam, A. 1973. The nucleotide sequence of the *lac* operator. Proc. Nat. Acad. Sci. USA 70:3581.

Grasso, R. J. and Buchanan, J. M. 1969. Synthesis of early RNA in bacteriophage T4-infected *Escherichia coli B*. Nature 224:882.

Grossman, L., Levine, S. S. and Allison, W. S. 1961. The reaction of formaldehyde with nucleotides and T2 bacteriophage DNA. J. Mol. Biol. 3:47.

Grossman, L. I., Cryer, D. R. Goldring, E. S. and Marmur, J. 1971. The petite mutation in yeast. J. Mol. Biol. 62:565.

Guha, A. and Szybalski, W. 1968. Fractionation of the complementary strands of coliphage T4 DNA based on the asymmetric distribution of the poly U and poly U,G binding sites. Virology 34:608.

Guha, A., Szybalski, W., Salser, W., Bolle, A., Geiduschek, E. P. and Pulitzer, J. F. 1971. Controls and polarity of transcription during bacteriophage T4 development. J. Mol. Biol. 59:329.

Hayward, S. D. and Smith, M. G. 1973. The chromosome of bacteriophage T5 III. Patterns of transcription from the single-stranded DNA fragments. J. Mol. Biol. 80:345.

Hinkle, D. C. and Chamberlin, J. J. 1972. Studies of the binding of Escherichia coli RNA polymerase to DNA. I. The role of sigma subunit site selection. J. Mol. Biol. 70:157.

Hinkle, D. C., Ring, J. and Chamberlin, M. 1972. Studies of the binding of Escherichia coli RNA polymerase to DNA III. Tight binding of RNA polymerase holoenzyme to single-strand breaks in T7 DNA. J. Mol. Biol. 70:197.

Hurwitz, J., Furth, J. J., Anders, M. and Evans, A. 1962. The role of deoxyribonucleic acid in ribonucleic acid synthesis. II. The influence of deoxyribonucleic acid on the reaction. J. Biol. Chem. 237:3752.

Josse, J., Kaiser, A. D. and Kornberg, A. 1961. Enzymatic synthesis of deoxyribonucleic acid VIII. Frequencies of nearest-neighbor base sequences in deoxyribonucleic acid. J. Biol. Chem. 236:864.

Kaiser, A. D. and Baldwin, R. L. 1962. A relation between dinucleotide and base frequencies in bacterial DNAs. J. Mol. Biol. 4:418.

Karkas, J., Rudner, R. and Chargaff, E. 1968. Separation of *B. subtilis* DNA into complementary strands, II. Template function and composition as determined by transcription with RNA polymerase. Proc. Nat. Acad. Sci. USA 60:915.

Karkas, J. D., Rudner, R. and Chargaff, E. 1970. Template properties of complementary fractions of denatured microbial deoxyribonucleic acids. Proc. Nat. Acad. Sci. USA 65:1049.

Kramer, R. A., Rosenberg, M. and Steitz, J. A. 1974. Nucleotide sequences of the 5' and 3' termini of bacteriophage T7 early messenger RNAs synthesized *in vivo*: evidence for sequence specificity in RNA processing. J. Mol. Biol. 89:767.

Langridge, R. 1969. Nucleic acids and polynucleotides. J. Cell. Physiol. 74: Suppl. 1, p 1.

Lanni, Y. T. and Szybalski, W. 1969. Transcription patterns for coliphage T5: hybridization between ^{3}H-RNAs and single-stranded T5$^+$ DNA fragments fractionated by the poly (U,G)-CsCl technique. Bacteriol. Proc. 69:192.

Loewen, P. C., Sekiya, T. and Khorana, H. G. 1974. The nucleotide sequence adjoining the CCA end of an *Escherichia coli* tyrosine transfer ribonucleic acid gene. J. Biol. Chem. 249:217.

Lowry, O., Rosenbrough, N. J., Farr, A. K. and Randall, R. S. 1951. Protein measurement with the Folin phenol reagent. J. Biol. Chem. 193:265.

Lunan, K. D. and Sinsheimer, R. L. 1956. A study of the nucleic acid of bacteriophage T7. Virology 2:455.

Maitra, U. and Hurwitz, J. 1965. The role of DNA in RNA synthesis IX. Nucleoside triphosphate termini in RNA polymerase products. Proc. Nat. Acad. Sci. USA 54:815.

Maniatis, T., Ptashne, M., Barrell, B. G. and Donelson, J. 1974. Sequence of a repressor binding site in the DNA of bacteriophage lambda. Nature 250:394.

Margulies, L., Remeza, V. and Rudner, R. 1970. Asymmetric template function of microbial deoxyribonucleic acids: transcription of ribosomal and soluble ribonucleic acids. J. Bact. 103:560.

Margulies, L., Remeza, V. and Rudner, R. 1971. Asymmetric template function of microbial deoxyribonucleic acids: transcription of messenger ribonucleic acid. J. Bact. 107:610.

Marmur, J. and Cordes, S. 1963. Studies on the complementary strands of bacteriophage DNA. Informational Macromolecules, A. J. Vogel et al., eds. Academic Press, New York p. 69.

Marmur, J. and Greenspan, C. M. 1963. Transcription in vitro of DNA from bacteriophage SP8. Science 142:387.

Maurer, R., Maniatis, T. and Ptashne, M. 1974. Promoters are in the operator in phage lambda. Nature 249:221.

Morgan, A. R. and Wells, R. D. 1968. Specificity of the three-stranded complex formation between double-stranded DNA and single-stranded RNA containing repeating nucleotide sequences. J. Mol. Biol. 37:63.

Mushynski, W. E. and Spencer, J. H. 1970a. Nucleotide clusters in deoxyribonucleic acids V. The pyrimidine oligonucleotides of strand r and l of bacteriophage T7 DNA. J. Mol. Biol. 52:91.

Mushynski, W. E. and Spencer, J. H. 1970b. Nucleotide clusters in deoxyribonucleic acids VI. The pyrimidine oligonucleotides of strands r and l of bacteriophage lambda DNA. J. Mol. Biol. 52:107.

Niyogi, S. K. 1969. The influence of chain length and base composition on the specific association of oligoribonucleotides with denatured deoxyribonucleic acid. J. Biol. Chem. 244:1576.

Niyogi, S. K. 1973. Annealing properties of oligoribonucleotides derived from T5 DNA-directed RNA compared to those from T2 and T7 DNA-directed RNAs. J. Biol. Chem. 218:2323.

Okamoto, T., Suguira, M. and Takanami, M. 1969. Length of RNA transcribed in the replicative form DNA of coliphage fd. J. Mol. Biol. 45:101.

Pispa, J. P. and Buchanan, J. M. 1971. Synthesis of bacteriophage T5 specific RNA _in vitro_. Biochem. Biophys. Acta 247:181.

Pribnow, D. 1975. Nucleotide sequence of an RNA polymerase site at an early T7 promoter. Proc. Nat. Acad. Sci. USA 72:784.

Richardson, C. C. 1966. The 5' terminal nucleotides of T7 bacteriophage deoxyribonucleic acid. J. Mol. Biol. 15:49.

Robertson, H. D. 1971. Enzymatic synthesis of bacteriophage fl DNA:RNA hybrid and double-stranded RNA. Nature 229:169.

Rubenstein, I. 1968. Heat-stable mutants of T5 phage I. The physical properties of the phage and their DNA molecules. Virology 36:356.

Rudner, R. and LeDoux, M. 1974. Distribution of pyrimidine oligonucleotides in complementary strand fractions of Escherichia coli deoxyribonucleic acid. Biochemistry 13:118.

Rudner, R., Karkas, J. D. and Chargaff, E. 1969. Separation of microbial deoxyribonucleic acids into complementary strands. Proc. Nat. Acad. Sci. 63:152.

Rudner, R., LeDoux, M. and Mazelis, A. 1972. Distribution of pyrimidine oligonucleotides in strands L and H of Bacillus subtilis DNA. Proc. Nat. Acad. Sci. USA 69:2745.

Salvo, R. A., Chakraboty, P. R. and Maitra, U. 1973. Studies on T3-induced ribonucleic acid polymerase IV. Transcription of denatured deoxyribonucleic acids preparations by T3 ribonucleic acid polymease. J. Biol. Chem. 248:6647.

Sambrook, J., Sharp, P. A. and Keller, W. 1972. Transcription of simian virus 40 I. Separation of the strands of SV40 DNA and hybridization of the separated strands to RNA extracted from lytically infected and transformed cells. J. Mol. Biol. 70:57.

Schaller, H., Gray, C. and Herrmann, K. 1975. Nucleotide sequence of an RNA polymerase binding site from the DNA of bacteriophage fd. Proc. Nat. Acad. Sci. USA 72:737.

Schaller, H., Voss, H. and Glucker, S. 1969. Structure of bacteriophage fd II. Isolation and characterization of a DNA fraction with double-strand-like properties. J. Mol. Biol. 44:445.

Sedat, J. W. and Sinsheimer, R. L. 1970. The *in vivo* ØX 174 mRNA. Cold Spring Harbor Symp. Quant. Biol. 35:163.

Sekiya, T. and Khorana, H. G. 1974. Nucleotide sequence in the promoter region of the *Escherichia coli* tyrosine tRNA gene. Proc. Nat. Acad. Sci. USA 71:2978.

Sheldrick, P. and Szybalski, W. 1969. Distribution of pyrimidine 'clusters' between the complementary DNA strands of certain *Bacillus* bacteriophage. J. Mol. Biol. 29:217.

Sinsheimer, R. L. 1968. Bacteriophage ØX 174 and related viruses. Prog. Nucl. Acid Res. Mol. Biol. 8:115.

Snedocor, Statistical Methods, Collegiate Press, Iowa State College, Ames, Iowa.

Sugimoto, K., Okamoto, T., Sugisaki, H. and Takanami, M. 1975. The nucleotide sequence of an RNA polymerase binding site on bacteriophage fd DNA. Nature 253:410.

Suguira, M., Okamoto, T. and Takanami, M. 1970. RNA polymerase and the selection of initiation site. Nature 225:598.

Summers, W. C. 1969. The process of infection with coliphage T7. 1. Characterization of T7 RNA by polyacrylimide gel electrophoretic analysis. Virology 39:175.

Summers, W. C. and Siegel, R. B. 1969. Control of specificity of *E. coli* RNA polymerase by a phage-coded protein. Nature 223:1111.

Summers, W. C. and Siegel, R. 1970. Regulation of coliphage T7 metabolism *in vivo* and *in vitro*. Cold Spring Harbor Symp. Quant. Biol. 35:253.

Summers, W. C. and Szybalski, W. 1968. Totally asymmetric transcription of coliphage T7 *in vivo*: correlation with poly G binding sites. Virology 34:9.

Swartz, M. N., Trautner, T. A. and Kornberg, A. 1962. Enzymatic synthesis of deoxyribonucleic acid XI. Further studies on nearest-neighbor base sequences in deoxyribonucleic acids. J. Biol. Chem. 237:1961.

Szybalski, W., Bovre, K., Fiandt, M. J., Guha, A., Hradecna, Z., Kumar, S., Lozeron, H. A., Maher, S. V. M., Nijkamp, H. J. J., Summers, W. C. and Tayloer, K. 1969. Transcriptional controls in developing bacteriophages. J. Cell. Physiol. 74:Suppl. 1, 33.

Szybalski, W., Kubinski, H. and Sheldrick, P. 1966. Pyrimidine clusters on the transcribing strand of DNA and their possible role in the initiation of RNA synthesis. Cold Spring Harbor Symp. Quant. Biol. 31:123.

Taylor, K., Hradecna, Z. and Szybalski, W. 1967. Asymmetric distribution of the transcribing regions on the complementary strands of phage lambda. Proc. Nat. Acad. Sci. USA 57:1618.

Tocchini-Valentini, G. P., Stodolsky, M., Aurisicchio, A., Sarnat, M., Graziosi, F., Weiss, S. D. and Geiduschek, E. P. 1963. On the asymmetry of RNA synthesis *in vivo*. Proc. Nat. Acad. Sci. USA 50:935.

Warner, R. C., Samuels, H. H., Abbott, M. T. and Krakow, J. S. 1963. Ribonucleic acid polymerase of *Azotobacter vinelandii*, II. Formation of DNA-RNA hybrids with single-stranded DNA as primer. Proc. Nat. Acad. Sci. USA 49:533.

Weiss, S. B. and Nakamoto, T. 1961. The enzymatic synthesis of RNA: nearest-neighbor base frequencies. Proc. Nat. Acad. Sci. USA 47:1400.

Wickner, W. and Kornberg, A. 1974. A novel form of RNA polymerase from *Escherichia coli*. Proc. Nat. Acad. Sci. USA 71:4425.

Weiss, S. B. and Nakamoto, T. 1961. The enzymatic synthesis of RNA: nearest-neighbor base frequencies. Proc. Nat. Acad. Sci. USA 47:1400.

Wilson, D. A. and Thomas, C. A., Jr. 1974. Palindromes in Chromosomes. J. Mol. Biol. 84:115.

Witmer, H. J. 1971. *In vitro* transcription of T4 deoxyribonucleic acid by *Escherichia coli* ribonucleic acid polymerase. J. Biol. Chem. 246:5220.

Wyatt, G. R. and Cohen, S. S. 1953. The bases of the nucleic acids of some bacterial and animal viruses: the occurrence of 5-hydroxymethylcytosine. Biochem. J. 55:774.

Table 1: Legend

a. % recovered refers to total T4 or dAT units recovered. The units recovered from the DEAE pooled enzyme are taken as 100% for dAT template.

b. Approximately 40% of the DEAE pooled enzyme was applied to the glycerol gradients. 13.4% dAT units represent a recovery of approximately 33% of the enzyme activity applied to the gradients.

Table 1: Purification of RNA polymerase from *E. coli*, according to Burgess *et al.* (1969) with some modifications

Enzyme Fraction	Protein		Units Recovered ($\times 10^3$)		Specific Activity (units/mg)		%Recovered[a]	
	Total mg	mg/ml	T4	dAT	T4	dAT	T4	dAT
I. Dnase treated extract	13,600	53.2	15.6	-	1.15	-	100	-
II: High speed supernatant	6.600	33.0	15.5	-	2.35	-	99	-
III: Ammonium sulfate fraction	1,780	11.9	16.3	-	9.2	-	104	-
DEAE pooled enzyme	165	3.3	9.3	100	56	632	60	100
Glycerol gradient pool	16.8	0.5	-	13.4	-	800	--	13.4[b]
Ammonium sulfate precipitated final fraction	7.8	2.6	3.0	6.5	390	840	19.2	6.5

Table 2: Legend

All transcripts were prepared using the standard enzyme reaction (Materials and Methods II). Reaction volume was 500 ul, except for T4 native DNA (both sets) and f1 RF DNA, which were 250 ul. T4 native DNA, limited-synthesis, was incubated for 6 minutes at 23 C.

The amount of protein and DNA are shown for one experiment, i.e. for one reaction mixture. Eight parallel experiments, two for each ^{32}P-labeled substrate were carried out. The DNA:protein ratios are derived from the saturation curves, as described in the text. To ensure a maximum amount of synthesis for the limited-synthesis T4 DNA transcription, the higher ratio of 1.0 was chosen for this reaction. The other T4 DNA transcripts were synthesized at a ratio of 1.0 also, in order to maintain the same conditions for all T4 DNA transcription.

a. $\% \text{ synthesis} = \dfrac{\text{nmoles of RNA synthesized}}{\text{nmoles DNA template available}}$

b. template efficiency =

$$\frac{\text{nmoles of RNA synthesized}}{\text{ug protein}}$$

Table 2: Optimum reaction conditions used in the preparation of ^{32}P-RNA transcripts for nearest-neighbor analysis: % synthesis, template efficiency

DNA template	Protein ug	DNA ug	DNA:protein ratio
T4 Native	6.2	6.2	1.0
T4 Native limited-synthesis	6.2	6.2	1.0
T4 Denatured	62	8.0	0.13
T5 Native	6.2	3.1	0.50
T5 Denatured	74	15	0.20
T7 Native	20	11	0.55
T7 Denatured	72	11	0.15
f1 RF	6.2	3.0	0.48

nmoles RNA made	DNA nmoles	% synthesis[a]	template[b] efficiency
5.4	20.4	26	0.87
0.75	20.4	3.6	0.12
6.0	26.4	23	0.10
5.1	10.2	50	0.82
7.7	50	15	0.10
27	36	75	1.4
12	36	33	0.17
4.4	9.9	45	0.71

Table 3: Legend

Analysis, by nuclease digestion, of products formed in the standard enzyme reaction, using native (N) and denatured (D) DNA templates, with *E. coli* RNA polymerase. Denatured T5 DNA is not included because there was not enough of the same T5 DNA preparation available to carry out these experiments.

After the standard enzyme reaction was completed, i.e. 20 minutes incubation at 37 C., aliquots were subjected to digestion with RNase (panRNase at 2 ug/ml and T1 RNase at 1 ug/ml), DNase (DNase I at 30 ug/ml) or both. Products of digested and undigested samples were then compared, as to the amount of labeled RNA that was acid precipitable and nitrocellulose membrane bound. Details of the experimental procedure are described in Material and Methods V.

The radioactivity present in each sample has been normalized to 100% for acid precipitable RNA in the undigested control sample.

Table 3: Analysis of transcription products by nuclease digestion: TCA precipitable, labeled RNA product (T), compared to nitrocellulose membrane bound product (M)

Nuclease treatment	Type of counts	T 4 DNA			
		N		D	
		%	cpm	%	cpm
None (control)	T	100	2724	100	1500
	M	14	388	91	1364
RNase	T	2.5	68	30	447
	M	2.2	60	39	579
DNase	T	50	1363	67	1003
	M	2.2	61	15	230
RNase and DNase	T	2.0	56	6.6	99
	M	0.8	21	4.2	63

T7 DNA				T5 DNA	
N		D		N	
%	cpm	%	cpm	%	cpm
100	5656	100	2239	100	4134
7.8	441	31	695	6.9	286
4.2	235	17	392	3.0	125
1.5	86	5.6	126	2.0	86
66	3706	89	1707	72	2984
0.8	50	16	351	1.1	44
0.4	25	5.4	122	2.4	99
0.3	18	3.6	82	0.2	10

Legend: Tables 4A, 4B and 4C

Table 4A: A sample set of data is given for the four 2'-(3') ribonucleoside monophosphates of one electrophoretic separation. The average fraction is the average of the four fractional values determined for the two aliquots of the duplicate reactions, that were run for each labeled substrate. The dinucleotide sequence associated with each transfer is also shown.

Table 4B: The base composition of the RNA transcripts was experimentally determined from the nearest-neighbor analysis. The base ratios and base composition for the f1 DNA transcripts are compared with the known values for f1 plus DNA.

Table 4C: Paired sequences I and II correspond to the six sets of dinucleotides with matching nearest-neighbor frequencies which are found in double-stranded DNA. Dinucleotides ApA, ApG, GpA and GpG are purine dinucleotides (NNF I). Dinucleotides UpU, CpU, UpC and CpC are pyrimidine dinucleotides (NNF II). The last two pairs are mixed purine-pyrimidine dinucleotides. The asymmetry ratio for each pair, i.e. NNF I/NNF II, has been calculated. Each pair has been evaluated for equivalence by the 't' test, and the p values are given.

a. s = standard error of the mean, and is equal to $D/n^{\frac{1}{2}}$, where D is the standard deviation and n is the number of entries.

Table 4A: Radioactivity measurements of ^{32}P-label transferred from 5'-nucleotides to 2'-(3')-nucleotides: RNA transcripts of f1 RF DNA

^{32}P-XTP	2'-(3') XMP	Sequence	cpm	Fraction.	Average Fraction
ATP	AMP	ApA	836	0.328	0.328
	GMP	GpA	492	0.193	0.196
	CMP	CpA	473	0.185	0.180
	UMP	UpA	750	0.294	0.295
			Sum 2551	1.000	1.000
GTP	AMP	ApG	558	0.199	0.205
	GMP	GpG	635	0.227	0.224
	CMP	CpG	590	0.211	0.205
	UMP	UpG	1017	0.363	0.365
			Sum 2800	1.000	1.000
CTP	AMP	ApC	430	0.198	0.196
	GMP	GpC	541	0.249	0.248
	CMP	CpC	465	0.214	0.216
	UMP	UpC	740	0.340	0.342
			Sum 2176	1.000	1.000
UTP	AMP	ApU	1369	0.235	0.232
	GMP	GpU	1087	0.187	0.180
	CMP	CpU	1229	0.211	0.213
	UMP	UpU	2136	0.367	0.374
			Sum 5821	1.000	1.000

Table 4B: Base composition and base ratios:
f1 (+) DNA; RNA transcripts of f1 RF DNA

Base composition (Mole %)

RNA transcripts		f1 (+) phage DNA*	
RNA base		DNA base	
A	0.242	A	0.245
G	0.202	G	0.202
C	0.205	C	0.205
U	0.346	T	0.348

Base Ratios

RNA transcripts		f1 (+) phage DNA	
RNA ratio		DNA ratio	
A/U	0.698	A/T	0.704
G/C	0.985	G/C	0.985
	Diasymmetry ratio		
A+U/G+C	1.44	A+T/G+C	1.46
	Purines/Pyrimidines		
A+G/U+C	0.806	A+G/T+C	0.810
	6-amino/6-keto		
A+C/U+G	0.815	A+C/T+G	0.818

* Schaller *et al.* (1969)

Table 4C: Nearest-neighbor frequencies (NNF) of RNA transcripts of double-stranded RF f1 DNA template

Sequence	NNF I	s $\times 10^2$	Sequence	NNF II	s $\times 10^2$	Asymmetry Ratio	p
ApA	0.0793	±0.01	UpU	0.130	±0.26	0.61	0.005
GpA	0.0475	±0.08	UpC	0.0701	±0.04	0.68	0.005
ApG	0.0423	±0.12	CpU	0.0737	±0.07	0.57	0.005
GpG	0.0462	±0.95	CpC	0.0441	±0.03	1.05	0.05
CpA	0.0438	±0.11	UpG	0.0752	±0.04	0.60	0.005
GpU	0.0622	±0.24	ApC	0.0401	±0.05	1.55	0.01
UpA	0.0714	±0.03					
ApU	0.0805	±0.09					
CpG	0.0422	±0.12					
GpC	0.0563	±0.02					

Legend: Tables 5A,B,C,D,E

Tables 5A,B,C: See Legend for Table 4A,B,C.

Table 5D: The nearest-neighbor frequencies of native and denatured DNA transcripts of each dinucleotide have been evaluated for equivalence by the 't' test, and the p values are given. From this comparison, changes in individual bases in the RNA transcripts of the native DNA are shown. The random nearest-neighbor frequencies have been calculated from the known DNA base composition of each phage studied (T7 DNA: Mushynski and Spencer, 1970 a; T4 and T5 DNAs: Wyatt and Cohen, 1953).

Table 5E: The distribution of dinucleotides in the RNA transcripts is shown for four groups of similar sequence.

Table 5A: Radioactivity measurements of ^{32}P-label transferred from 5'-nucleotides to 2'-(3')-nucleotides: RNA transcripts of T7 DNA

^{32}P-XTP	2'-(3') XMP	Native T7 DNA Sequence	cpm	Fraction	Average Fraction
ATP	AMP	ApA	1384	0.296	0.288
	GMP	GpA	1230	0.263	0.263
	CMP	CpA	1165	0.249	0.244
	UMP	UpA	902	0.193	0.205
			Sum 4681	1.000	1.000
GTP	AMP	ApG	1410	0.256	0.256
	GMP	GpG	1286	0.234	0.234
	CMP	CpG	1285	0.234	0.224
	UMP	UpG	1518	0.276	0.286
			Sum 5499	1.000	1.000
CTP	AMP	ApC	1031	0.306	0.302
	GMP	GpC	944	0.283	0.283
	CMP	CpC	565	0.169	0.172
	UMP	UpC	777	0.234	0.242
			Sum 3317	1.000	1.000
UTP	AMP	ApU	1198	0.238	0.238
	GMP	GpU	1186	0.236	0.238
	CMP	CpU	1542	0.307	0.297
	UMP	UpU	1103	0.219	0.226
			Sum 5029	1.000	1.000

^{32}P-XTP	2'-(3')-XMP	Denatured T7 DNA		Fraction	Average Fraction
		Sequence	cpm		
ATP	AMP	ApA	643	0.254	0.255
	GMP	GpA	670	0.265	0.261
	CMP	CpA	639	0.252	0.259
	UMP	UpA	583	0.230	0.226
			Sum 2535	1.000	1.000
GTP	AMP	ApG	874	0.276	0.284
	GMP	GpG	714	0.226	0.216
	CMP	CpG	701	0.222	0.220
	UMP	UpG	871	0.276	0.282
			Sum 3160	1.000	1.000
CTP	AMP	ApC	462	0.289	0.284
	GMP	GpC	336	0.210	0.227
	CMP	CpC	361	0.226	0.217
	UMP	UpC	440	0.275	0.271
			Sum 1599	1.000	1.000
UTP	AMP	ApU	666	0.223	0.223
	GMP	GpU	711	0.238	0.243
	CMP	CpU	830	0.278	0.282
	UMP	UpU	771	0.260	0.252
			Sum 2984	1.000	1.000

Table 5B: Base composition and base ratios: T7 DNA; RNA transcripts of T7 DNA

Base Composition (Mole %)

RNA transcripts			T7 phage DNA*	
RNA base	DNA template		DNA base	
	Native	Denatured		
A	0.279	0.262	T	0.253
G	0.234	0.237	C	0.249
C	0.241	0.245	G	0.237
U	0.247	0.257	A	0.260

Base Ratios

RNA transcripts			T7 phage DNA*	
RNA ratio	DNA template		DNA ratio	
	Native	Denatured		
A/U	1.13	1.02	A/T	0.97
G/C	0.97	0.97	G/C	1.05
		Dissymmetry ratio		
A+U/G+C	1.11	1.08	A+T/G+C	1.06
		Purines/Pyrimidines		
A+G/U+C	1.05	1.00	A+G/T+C	0.98
		6-amino/6-keto		
A+C/U+C	1.08	1.03	A+C/T+C	1.04

* Mushynski and Spencer, (1970 a)

Table 5C: Nearest-neighbor frequencies NNF of NA transcripts of native and denatured T7 DNA templates

Native T7 DNA template

Sequence	NNF I	s $x10^2$	Sequence	NNF II	s $x10^2$	Asymmetry Ratio	p
ApA	0.0804	±0.05	UpU	0.0684	±0.06	1.18	0.001
GpA	0.0734	±0.06	UpC	0.0583	±0.05	1.26	0.001
ApG	0.0599	±0.04	CpU	0.0734	±0.07	0.82	0.001
GpG	0.0545	±0.04	CpC	0.0414	±0.02	1.32	0.001
CpA	0.0681	±0.04	UpG	0.0669	±0.10	1.02	0.30
GpU	0.0590	±0.03	ApC	0.0728	±0.07	0.80	0.001
UpA	0.0575	±0.11					
ApU	0.0588	±0.02					
CpG	0.0524	±0.06					
GpA	0.0682	±0.04					

Denatured T7 DNA template

Sequence	NNF I	s $\times 10^2$	Sequence	NNF II	s $\times 10^2$	Asymmetry Ratio	p
ApA	0.0668	±0.04	UpU	0.0650	±0.10	1.02	0.20
GpA	0.0683	±0.06	UpC	0.0664	±0.07	1.03	0.10
ApG	0.0671	±0.07	CpU	0.0728	±0.06	0.92	0.001
GpG	0.0510	±0.12	CpC	0.0532	±0.09	0.96	0.20
CpA	0.0678	±0.10	UpG	0.0668	±0.15	1.01	0.50
GpU	0.0627	±0.05	ApC	0.0696	±0.08	0.90	0.001
UpA	0.0592	±0.14					
ApU	0.0575	±0.02					
CpG	0.0521	±0.14					
GpC	0.0556	±0.14					

Table 5D: Comparison of dinucleotide frequencies of native (N) and denatured (D) T7 RNA transcripts: frequencies calculated for random association of bases

Sequence	NNF RNA transcripts		N/D	p	Base change in N + increase - decrease 0 no change
	N	D			
ApA	0.0804	0.0668	1.20	0.01	+ A
UpU	0.0684	0.0650	1.05	0.001	- U
GpA	0.0734	0.0683	1.08	0.001	+ G.A
UpC	0.0583	0.0664	0.88	0.001	- U,C
ApG	0.0599	0.0671	0.89	0.001	- A,G
CpU	0.0734	0.0728	1.01	0.001	+ U,C
GpG	0.0545	0.0510	1.07	0.001	+ G
CpC	0.0414	0.0532	0.78	0.01	- C
CpA	0.0681	0.0678	1.00	0.50	0
UpG	0.0669	0.0668	1.00	0.50	0
GpU	0.0590	0.0627	0.94	0.01	- G,U
ApC	0.0728	0.0696	1.05	0.001	+ A,C
UpA	0.0575	0.0592	0.97	0.40	0
ApU	0.0588	0.0575	1.02	0.001	+ A,U
CpG	0.0524	0.0521	1.01	0.50	0
GpC	0.0682	0.0556	1.23	0.001	+ G,C

Sequence	Random NNF
ApA	0.0676
UpU	0.0640
GpA	0.0616
UpC	0.0630
ApG	0.0616
CpU	0.0630
GpG	0.0562
CpC	0.0620
CpA	0.0645
UpG	0.0600
GpU	0.0600
ApC	0.0645
UpA	0.0658
ApU	0.0658
CpG	0.0590
GpC	0.0590

Table 5E: Dinucleotide distribution of RNA transcripts of T7 DNA

	Dinucleotide groups*			
	I	II	III	IV
	ApA	UpU	CpA	UpA
	ApG	CpU	UpG	ApU
	GpA	UpC	ApC	CpG
	GpG	CpC	GpU	GpC
RNA		Mole %		
RNA transcript of native T7 DNA	0.268	0.242	0.267	0.237
RNA transcript of denatured T7 DNA	0.253	0.257	0.267	0.224

* groups are:

I. Purine dinucleotides

II. Pyrimidine dinucleotides

III. Mixed complementary dinucleotides

IV. Mixed self-complementary dinucleotides

Legend: Tables 6A,B,C,D,E

See Legend for Tables 4A,B,C and 5D,E. In Table 6D, the nearest-neighbor frequencies of the T4 phage DNA are shown. In Table 6E, the distribution of the dinucleotides has also been given for T4 phage DNA.

Table 6A: Radioactivity measurements of ^{32}P-label transferred from 5'-nucleotides to 2'Q(3')-nucleotides; RNA transcripts of T4 DNA

		Native T4 DNA			
^{32}P-XTP	2'-(3') XMP	Sequence	cpm	Fraction	Average Fraction
ATP	AMP	ApA	4512	0.381	0.375
	GMP	GpA	2464	0.208	0.197
	CMP	CpA	1766	0.149	0.156
	UMP	UpA	3087	0.261	0.271
		Sum	11829	1.000	1.000
GTP	AMP	ApG	1860	0.275	0.279
	GMP	GpG	1245	0.185	0.179
	CMP	CpG	1147	0.170	0.158
	UMP	UpG	2490	0.369	0.383
		Sum	6742	1.000	1.000
CTP	AMP	ApC	1091	0.270	0.265
	GMP	GpC	851	0.211	0.222
	CMP	CpC	668	0.165	0.160
	UMP	UpC	1431	0.354	0.352
		Sum	4041	1.000	1.000
UTP	AMP	ApU	1772	0.295	0.308
	GMP	GpU	801	0.133	0.145
	CMP	CpU	1061	0.176	0.169
	UMP	UpU	2380	0.396	0.376
		Sum	6014	1.000	1.000

^{32}P-XTP	2'-(3') XMP	Denatured T4 DNA		Fraction	Average Fraction
		Sequence	cpm		
ATP	AMP	ApA	1647	0.380	0.380
	GMP	GpA	851	0.196	0.205
	CMP	CpA	665	0.154	0.155
	UMP	UpA	1168	0.290	0.260
			Sum 4331	1.000	1.000
GTP	AMP	ApG	716	0.284	0.292
	GMP	GpG	459	0.182	0.183
	CMP	CpG	385	0.153	0.152
	UMP	UpG	960	0.381	0.372
			Sum 2520	1.000	1.000
CTP	AMP	ApC	530	0.273	0.280
	GMP	GpC	411	0.212	0.209
	CMP	CpC	305	0.157	0.152
	UMP	UpC	696	0.358	0.356
			Sum 1902	1.000	1,000
UTP	AMP	ApU	2023	0.326	0.320
	GMP	GpU	1043	0.168	0.164
	CMP	CpU	985	0.159	0.162
	UMP	UpU	2151	0.347	0.354
			Sum 6202	1.000	1.000

Table 6B: Base composition and base ratios:
T4 DNA; RNA transcripts of T4 DNA

Base Composition (Mole %)

RNA transcripts			T4 phage DNA*	
RNA base	DNA template		DNA base	
	Native	Denatured		
A	0.317	0.328	T	0.318
G	0.181	0.188	C	0.179
C	0.161	0.157	G	0.184
U	0.340	0.328	A	0.319

Base ratios

RNA transcripts			T4 phage DNA*	
RNA ratio	DNA template		DNA ratio	
	Native	Denatured		
A/U	0.93	1.00	A/T	1.00
G/C	1.13	1.20	G/C	0.97
		Dissymmetry ratio		
A+U/G+C	1.92	1.90	A+T/G+C	1.75
		Purines/Pyrimidines		
A+G/U+C	0.99	1.06	A+G/T+C	0.99
		6-amino/6-keto		
A+C/U+G	0.92	0.94	A+C/T+G	0.99

* Wyatt and Cohen, (1953)

Table 6C: Nearest-neighbor frequencies (NNF) of RNA transcripts of native and denatured T4 DNA templates

Native T4 DNA template

Sequence	NNF I	s x10^2	Sequence	NNF II	s x10^2	Asymmetry Ratio	p
ApA	0.119	±0.13	UpU	0.128	±0.04	0.93	0.001
GpA	0.0625	±0.30	UpC	0.0567	±0.06	1.10	0.10
ApG	0.0506	±0.08	CpU	0.0574	±0.15	0.88	0.10
GpG	0.0325	±0.03	CpC	0.0258	±0.05	1.26	0.001
CpA	0.0496	±0.11	UpG	0.0693	±0.11	0.72	0.001
GpU	0.0498	±0.34	ApC	0.0426	±0.11	1.17	0.10
UpA	0.0858	±0.04					
ApU	0.105	±0.22					
CpG	0.0286	±0.08					
GpC	0.0358	±0.16					

Denatured T4 DNA template

Sequence	NNF I	s $\times 10^2$	Sequence	NNF II	s $\times 10^2$	Asymmetry Ratio	p
ApA	0.125	±0.02	UpU	0.116	±0.10	1.08	0.001
GpA	0.0673	±0.07	UpC	0.0558	±0.03	1.20	0.001
ApG	0.0530	±0.07	CpU	0.0529	±0.03	1.00	0.50
GpG	0.0344	±0.04	CpC	0.0239	±0.03	1.44	0.001
CpA	0.0510	±0.01	UpG	0.0699	±0.01	0.73	0.001
GpU	0.0534	±0.06	ApC	0.0439	±0.04	1.22	0.001
UpA	0.0854	±0.08					
ApU	0.105	±0.08					
CpG	0.0286	±0.02					
GpC	0.0329	±0.02					

Table 6D: Comparison of dinucleotide frequencies of native (N) and denatured (D) T4 RNA transcripts: dinucleotide frequencies for random association of bases and for T4 phage DNA

Sequence	NNF RNA transcripts		N/D	p	Base change in N + increase - decrease 0 no change
	N	D			
ApA	0.119	0.125	0.95	0.001	- A
UpU	0.128	0.116	1.11	0.001	+ U
GpA	0.0625	0.0673	0 93	0.05	- G,A
UpC	0.0567	0.0558	1.02	0.20	0
ApG	0.0506	0.0530	0.95	0.05	- A,G
CpU	0.0574	0.0529	1.08	0.005	+ C,U
GpG	0.0325	0.0344	0.94	0.05	- G
CpC	0.0258	0.0239	1.08	0.001	+ C
CpA	0.0496	0.0510	0.97	0.05	- A,C
UpG	0.0693	0.0699	0.99	0.50	0
GpU	0.0498	0.0534	0.93	0.20	0
ApC	0.0426	0.0439	0.97	0.20	0
UpA	0.0858	0.0854	1.00	0.50	0
ApU	0.105	0.105	1.00	0.50	0
GpG	0.0286	0.0286	1.00	0.50	0
GpC	0.0358	0.0329	1.09	0.01	+ G,C

Sequence RNA	Random NNF	T4 DNA*	Sequence DNA
ApA	0.101	0.109	TpT
UpU	0.102	0.109	ApA
GpA	0.569	0.059	TpC
UpC	0.0587	0.059	GpA
ApG	0.0569	0.056	CpT
CpU	0.0587	0.057	ApG
GpG	0.0320	0.032	CpC
CpC	0.0338	0.035	GpG
CpA	0.0585	0.063	TpG
UpG	0.0571	0.061	CpA
GpU	0.0571	0.049	ApC
ApC	0.0585	0.052	GpT
UpA	0.101	0.091	TpA
ApU	0.101	0.105	ApT
CpG	0.0329	0.027	CpG
GpC	0.0329	0.036	GpC

* from Josse *et al.* (1961)

Table 6E: Dinucleotide distribution of RNA transcripts of T4 DNA

	Dinucleotide groups*			
	I	II	III	IV
	ApA ApG GpA GpG	UpU CpU UpC CpC	CpA UpG ApC GpU	UpA ApU CpG GpC
RNA or DNA		Mole %		
RNA transcript of native T4 DNA	0.294	0.245	0.214	0.247
RNA transcript of denatured T4 DNA	0.280	0.249	0.218	0.252
T4 phage DNA #	0.260	0.256	0.225	0.259

* groups are: I. Purine dinucleotides

II. Pyrimidine dinucleotides

III. Mixed complementary dinucleotides

IV. Mixed self-complementary dinucleotides

\# Josse *et al*. (1961)

Legend: Tables 7A,B,C,D,E

See Legend for Tables 4A,B,C and 5D,E. In Table 7D, the nearest-neighbor frequencies of the T4 phage DNA are shown. In Table 7E, the distribution of the dinucleotides has also been given for T4 phage DNA.

Table 7A: Radioactivity measurements of ^{32}P-label transferred from 5'-nucleotides to 2'-(3')-nucleotides: limited-synthesis transcripts of native T4 DNA

^{32}P-XTP	2'-(3') XTP	Sequence	cpm	Fraction	Average Fraction
ATP	AMP	ApA	1131	0.384	0.388
	GMP	GpA	641	0.217	0.222
	CMP	CpA	421	0.143	0.135
	UMP	UpA	753	0.256	0.254
			Sum 2946	1.000	1.000
GTP	AMP	ApG	418	0.303	0.302
	GMP	GpG	256	0.186	0.174
	CMP	CpG	159	0.115	0.126
	UMP	UpG	545	0.396	0.398
			Sum 1378	1.000	1.000
CTP	AMP	ApC	321	0.266	0.263
	GMP	GpC	290	0.240	0.235
	CMP	CpC	167	0.138	0.139
	UMP	UpC	430	0.356	0.362
			Sum 1208	1.000	1.000
UTP	AMP	ApU	528	0.324	0.315
	GMP	GpU	246	0.151	0.164
	CMP	CpU	251	0.154	0.156
	UMP	UpU	603	0.370	0.364
			Sum 1628	1.000	1.000

Table 7B: Base composition and base ratios:
T4 DNA; limited-synthesis RNA transcripts of T4 DNA

Base Composition (mole %)

RNA transcripts			T4 phage DNA*	
RNA base	DNA template		DNA base	
	Native	Denatured		
A	0.321	0.328	T	0.318
G	0.185	0.188	C	0.179
C	0.167	0.157	G	0.184
U	0.340	0.328	A	0.319

Base ratios

RNA transcripts			T4 phage DNA*	
RNA ratio	DNA template		DNA ratio	
	Native	Denatured		
A/U	0.98	1.00	A/T	1.00
G/C	1.11	1.20	G/C	0.97
		Dissymmetry ratio		
A+U/G+C	1.84	1.90	A+T/G+C	1.75
		Purines/Pyrimidines		
A+G/U+C	0.98	1.06	A+G/T+C	0.99
		6-amino/6-keto		
A+C/U+G	0.95	0.94	A+C/T+C	0.99

* Wyatt and Cohen (1953)

Table 7C: Nearest-neighbor frequencies (NNF) of RNA transcripts of native T4 DNA (limited-synthesis)

Sequence	NNF I	s x10^2	Sequence	NNF II	s x10^2	Asymmetry Ratio	p
ApA	0.128	±0.07	UpU	0.122	±0.41	1.05	0.20
GpA	0.0731	±0.11	UpC	0.0511	±0.05	1.43	0.001
ApG	0.0588	±0.11	CpU	0.0524	±0.18	1.12	0.02
GpG	0.0340	±0.08	CpC	0.0197	±0.02	1.73	0.001
CpA	0.0445	±0.09	UpG	0.0777	±0.19	0.52	0.001
GpU	0.0548	±0.49	ApC	0.0372	±0.07	1.48	0.02
UpA	0.0838	±0.06					
ApU	0.105	±0.19					
CpG	0.0246	±0.12					
GpC	0.0332	±0.07					

Table 7D: Comparison of dinucleotide frequencies of limited synthesis native (N) and denatured (D) T4 RNA transcripts: dinucleotide frequencies for random association of bases and for T4 phage DNA

Sequence	NNF		N/D	p	Base change in N + increase - decrease 0 no change
	RNA transcripts				
	N	D			
ApA	0.128	0.125	1.02	0.30	0
UpU	0.122	0.116	1.06	0.05	+ U
GpA	0.0731	0.0673	1.09	0.001	+ G,A
UpC	0.0511	0.0558	0.92	0.001	- U,C
ApG	0.0588	0.0530	1.11	0.001	+ A,G
CpU	0.0524	0.0529	0.99	0.50	0
GpG	0.0340	0.0344	0.99	0.50	0
CpC	0.0197	0.0239	0.82	0.001	- C
CpA	0.0445	0.0510	0.87	0.30	0
UpC	0.0777	0.0699	1.11	0.001	+ G,U
GpU	0.0548	0.0534	1.03	0.50	0
ApC	0.0372	0.0439	0.85	0.001	- A,C
UpA	0.0838	0.0854	0.98	0.30	0
ApU	0.105	0.105	1.00	0.50	0
CpG	0.0246	0.0286	0.86	0.005	- G,C
GpC	0.0332	0.0329	1.01	0.50	0

Sequence RNA	Random NNF	T4 DNA*	Sequence DNA
ApA	0.101	0.109	TpT
UpU	0.102	0.109	ApA
GpA	0.0569	0.059	TpC
UpC	0.0587	0.059	GpA
ApG	0.0569	0.056	CpT
CpU	0.0587	0.057	ApG
GpG	0.0320	0.032	CpC
CpC	0.0338	0.035	GpG
CpA	0.0585	0.063	TpG
UpG	0.0571	0.061	CpA
GpU	0.0571	0.049	ApC
ApC	0.0585	0.052	GpT
UpA	0.101	0.091	TpA
ApU	0.101	0.105	ApT
CpG	0.0329	0.027	CpG
GpC	0.0329	0.036	GpC

* from Josse *et al*. (1961)

Table 7E: Dinucleotide distribution of RNA transcripts of T4 DNA (limited-synthesis)

	Dinucleotide groups*			
	I	II	III	IV
	ApA	UpU	CpA	UpA
	ApG	CpU	UpG	ApU
	GpA	UpC	ApC	CpG
	GpG	CpC	GpU	GpC
RNA or DNA	Mole %			
RNA transcripts of native DNA (limited synthesis)	0.294	0.245	0.214	0.247
RNA transcripts of denatured T4 DNA	0.280	0.249	0.218	0.252
T4 phage DNA #	0.260	0.256	0.225	0.259

* groups are: I. Purine dinucleotides
II. Pyrimidine dinucleotides
III. Mixed complementary dinucleotides
IV. Mixed self-complementary dinucleotides

Josse *et al*. (1961)

Legend: Tables 8A,B,C,D,E

See Legend for Tables 4A,B,C and Tables 5D,E. In Table 8D, the nearest-neighbor frequencies of the T5 phage DNA are shown. In Table 7E, the distribution of the dinucleotides has also been given for T5 phage DNA.

Table 8A: Radioactivity measurements of ^{32}P-label transferred from 5'-nucleotides to 2'-(3')-nucleotides: RNA transcripts of T5 DNA

		Native T5 DNA			
^{32}P-XTP	2'-(3') XMP	Sequence	cpm	Fraction	Average Fraction
ATP	AMP	ApA	2386	0.341	0.334
	GMP	GpA	1417	0.202	0.202
	CMP	CpA	1100	0.157	0.163
	UMP	UpA	2099	0.300	0.301
		Sum	7002	1.000	1.000
GTP	AMP	ApG	1309	0.308	0.311
	GMP	GpG	897	0.211	0.210
	CMP	CpG	705	0.166	0.162
	UMP	UpG	1334	0.314	0.318
		Sum	4245	1.000	1.000
CTP	AMP	ApC	1044	01272	01274
	GMP	GpC	903	0.236	0.234
	CMP	CpC	740	0.193	0.198
	UMP	UpC	1148	0.299	0.294
		Sum	3835	1.000	1.000
UTP	AMP	ApU	1661	0.294	0.292
	GMP	GpU	1084	0.192	0.182
	CMP	CpU	1130	0.200	0.206
	UMP	UpU	1783	0.315	0.320
		Sum	5656	1.000	1.000

^{32}P-XTP	2'-(3') XMP	Denatured T5 DNA Sequence	cpm	Fraction	Average Fraction
ATP	AMP	ApA	1468	0.340	0.334
	GMP	GpA	810	0.189	0.184
	CMP	CpA	756	0.176	0.177
	UMP	UpA	1264	0.295	0.304
		Sum	4285	1.000	1.000
GTP	AMP	ApG	1187	0.300	0.302
	GMP	GpG	761	0.193	0.202
	CMP	CpG	709	0.179	0.173
	UMP	UpG	1295	0.328	0.322
		Sum	3952	1.000	1.000
CTP	AMP	ApC	776	0.270	0.263
	GMP	GpC	683	0.238	0.222
	CMP	CpC	553	0.192	0.204
	UMP	UpC	862	0.300	0.311
		Sum	2874	1.000	1.000
UTP	AMP	ApU	1511	0.276	0.280
	GMP	GpU	966	0.176	0.174
	CMP	CpU	1153	0.210	0.206
	UMP	UpU	1852	0.338	0.340
		Sum	5477	1.000	1.000

Table 8B: Base composition and base ratios: T5 DNA; RNA transcripts of T5 DNA

Base Composition (Mole %)

RNA transcripts			T5 phage DNA*	
RNA base	DNA template		DNA base	
	Native	Denatured		
A	0.304	0.284	T	0.303
G	0.203	0.225	C	0.195
C	0.188	0.183	G	0.195
U	0.311	0.308	A	0.308

Base ratios

RNA transcripts			T5 phage DNA*	
RNA ratio	DNA template		DNA ratio	
	Native	Denatured		
A/U	0.98	1.08	A/T	0.98
G/C	1.08	1.23	G/C	1.00
		Dissymmetry ratios		
A+U/G+C	1.57	1.45	A+T/G+C	1.57
		Purines/Pyrimidines		
A+G/U+C	1.02	1.04	A+G/T+C	1.01
		6-amino/6-keto		
A+C/U+G	0.96	0.88	A+C/T+C	0.99

* Wyatt and Cohen (1953)

Table 8C: Nearest-neighbor frequencies (NNF) of RNA transcripts of native and denatured T5 DNA templates

Native T5 DNA template

Sequence	NNF I	s x10^2	Sequence	NNF II	s x10^2	Asymmetry Ratio	p
ApA	0.102	±0.12	UpU	0.0998	±0.14	1.02	0.20
GpA	0.0606	±0.10	UpC	0.0551	±0.08	1.11	0.001
ApG	0.0632	±0.04	CpU	0.0640	±0.08	0.99	0.40
GpG	0.0426	±0.02	CpC	0.0371	±0.03	1.15	0.001
CpA	0.0495	±0.04	UpG	0.0644	±0.05	0.77	0.001
GpU	0.0567	±0.10	ApC	0.0514	±0.06	1.10	0.001
UpA	0.0917	±0.07					
ApU	0.0907	±0.11					
CpG	0.0328	±0.02					
GpC	0.0440	±0.02					

Denatured T5 DNA template

Sequence	NNF I	s $\times 10^2$	Sequence	NNF II	s $\times 10^2$	Asymmetry Ratio	p
ApA	0.0950	±0.17	UpU	0.105	±0.24	0.90	0.10
GpA	0.0524	±0.12	UpC	0.0570	±0.07	0.92	0.005
ApG	0.0680	±0.03	CpU	0.0633	±0.16	1.07	0.05
GpG	0.0455	±0.08	CpC	0.0374	±0.13	1.22	0.001
CpA	0.0502	±0.02	UpG	0.0725	±0.09	0.69	0.001
GpU	0.0536	±0.05	ApC	0.0481	±0.07	1.11	0.001
UpA	0.0864	±0.26					
ApU	0.0864	±0.09					
CpG	0.0390	±0.07					
GpC	0.0406	±0.16					

Table 8D: Comparison of dinucleotide frequencies of native (N) and denatured (D) T5 RNA transcripts: dinucleotide frequencies for random association of bases and for T5 phage DNA

Sequence	NNF RNA transcripts		N/D	p	Base change in N + increase - decrease 0 no change
	N	D			
ApA	0.102	0.0950	1.07	0.025	+ A
UpU	0.099	0,105	0.95	0.05	- U
GpA	0.0606	0.0524	1.20	0.005	+ G,A
UpC	0.0551	0.0570	0.97	0.10	0
ApG	0.0632	0.0680	0.93	0;001	- A,G
CpU	0.0640	0.0633	1.01	0.50	0
GpG	0.0426	0.0455	0.94	0.001	- G
CpC	0.0371	0.0374	0.99	0.50	0
CpA	0.0495	0.0502	0.99	0.40	0
UpG	0.0644	0.0725	0.89	0.001	- G,U
GpU	0.0567	0.0536	1.06	0.10	0
ApC	0.0514	0.0481	1.07	0.005	+ A,C
UpA	0.0917	0.0864	1.06	0.025	+ A,U
ApU	0.0907	0.0864	1.05	0.05	+ A,U
CpG	0.0328	0.0390	0.84	0.10	0
GpC	0.0440	0.0406	1.08	0.01	+ G,C

Sequence RNA	Random NNF	T5 DNA*	Sequence DNA
ApA	0.0949	0.100	TpT
UpU	0.0919	0.105	ApA
GpA	0.0601	0.06	TpC
UpC	0.0591	0.054	GpA
ApG	0.0601	0.064	CpT
CpU	0.0591	0.056	ApG
GpG	0.0380	0.038	CpC
CpC	0.0380	0.035	GpG
CpA	0.0601	0.057	TpG
UpG	0.0591	0.058	CpA
GpU	0.0591	0.052	ApC
ApC	0.0601	0.048	GpT
UpA	0.0933	0.098	TpA
ApU	0.0933	0.103	ApT
CpG	0.0380	0.032	CpG
GpC	0.0380	0.043	GpC

* from Swartz _et al_. (1962)

Table 8E: Dinucleotide distribution of RNA transcripts of T5 DNA

	Dinucleotide groups*			
	I	II	III	IV
	ApA	UpU	CpA	UpA
	ApG	CpU	UpG	ApU
	GpA	UpC	GpU	CpG
	GpG	CpC	ApC	GpC
RNA or DNA	Mole %			
RNA transcripts of native T5 DNA	0.268	0.256	0.202	0.259
RNA transcripts of denatured T5 DNA	0.261	0.263	0.224	0.252
T5 phage DNA #	0.250	0.262	0.215	0.276

* groups are:
I. Purine dinucleotides
II. Pyrimidine dinucleotides
III. Mixed complementary dinucleotides
IV. Mixed self-complementary dinucleotides

Swartz _et al_. (1962)

Legend: Table 9

The variances in this table have been calculated for the first four pairs of dinucleotides, and also for all six dinucleotide pairs. The four dinucleotide pairs are purine;pyrimidine pairs:

ApA:UpU, ApG:CpU, GpA:UpC, GpG:CpC;

The six dinucleotide pairs include these four,and the two mixed purine-pyrimidine pairs:

CpA:UpG and GpU:ApC.

The values of the variances are:

$$\text{Observed} = \frac{\Sigma(\text{Asymmetry ratio predicted-observed asymmetry ratio})^2}{n - 1}$$

$$\text{Error} = \frac{\Sigma(v \text{ x Asymmetry ratio predicted})^2}{n - 1}$$

where n = number of entried, v = coefficient of variation, see Appendix B.

The asymmetry ratio is discussed in the text, and the values determined for all the DNAs studied are given in Tables 5C, 6C, 7C and 8C. The observed asymmetry ratio is defined as the ratio of the nearest-neighbor frequencies, in the transcribed RNA, of the matching dinucleotide pairs. The predicted asymmetry ratio is defined as 1.0, its value in double-stranded DNA.

a. T4 Native (limited) refers to the experiment, discussed in the text, in which RNA synthesis was limited to early mRNA.

Table 9: Variances of the asymmetry ratios of the paired dinucleotides in the RNA transcripts

DNA template	Variances $(\times 10^2)$					
	4 Dinucleotide pairs			6 Dinucleotide pairs		
	Observed	Error	p	Observed	Error	p
T4 Native	3.2	0.24	< 0.05	4.1	0.29	< 0.05
T4 Denatured	7.9	0.16	< 0.05	7.2	0.12	< 0.05
T4 Native (limited)	24	0.33	< 0.05	24	0.59	< 0.05
T4 Denatured	7.9	0.16	< 0.05	7.2	0.12	< 0.05
T5 Native	1.2	0.16	< 0.05	1.2	0.15	< 0.05
T5 Denatured	2.3	0.20	< 0.05	3.6	0.13	< 0.05
T7 Native	7.8	0.09	< 0.05	5.5	0.08	< 0.05
T7 Denatured	0.31	0.10	0.20	0.39	0.09	0.10

Legend: Table 10

The sixteen dinucleotide frequencies found experimentally, for each phage DNA, have been compared to the random nearest-neighbor frequencies of these dinucleotides. The random values used, are given in Tables 5D, 6D, 7D and 8D. The variance of the observed frequencies (NNF observed) from the predicted random frequencies (NNF random) was calculated, as follows:

$$\text{Variance observed} = \frac{\Sigma(\text{NNF observed} - \text{NNF random})^2}{n - 1}$$

The variance of the predicted values, calculated from experimental error, was also determined, as follows:

$$\text{Variance error} = \frac{\Sigma(v \times \text{NNF predicted})^2}{n - 1}$$

where n = number of entries, and v = coefficient of variation (see Appendix B).

Table 10: Variance from random nearest-neighbor frequencies for each bacteriophage DNA

DNA template	Variance ($x 10^4$)		
	Observed	Error	p
T4 Native	1.3	0.10	0.001
T4 Native (limited)	2.1	0.12	0.001
T4 Denatured	1.2	0.03	0.001
T5 Native	0.32	0.06	0.001
T5 Denatured	0.62	0.05	0.001
T7 Native	0.84	0.03	0.001
T7 Denatured	0.36	0.03	0.001

Legend: Table 11

The experimental dinucleotide frequencies have been compared to the calculated random frequencies (Tables 5D, 6D, 7D and 8D) for each dinucleotide sequence. The variance has been determined for the transcripts of the four native DNA tempäates and the three denatured DNA templates, combined (native: T4, T4-limited, T5 and T7; denatured: T4, T5 and T7). The calculations are described in Table 10 and in Appendix B.

Table 11: Variance from random nearest-neighbor frequencies for each dinucleotide sequence

Dinucleotide	Variance (x 10^4)			Deviation from random frequencies ± 0.005		
	Observed	Error	p	+	0	-
ApA	3.2	0.026	< 0.05	5	2	
UpU	2.5	0.15	< 0.05	5	2	
GpA	1.06	0.074	< 0.05	5	1	1
UpC	0.21	0.024	< 0.05		6	1
ApG	0.27	0.041	< 0.05	2	4	1
GpG	0.20	0.020	< 0.05	1	5	1
CpC	1.44	0.030	< 0.05		2	5
CpA	0.94	0.030	< 0.05		2	5
UpG	1.73	0.053	< 0.05	7		
GpU	0.20	0.26	> 0.05		5	2
ApC	2.0	0.18	< 0.05	1	1	5
UpA	1.55	0.087	< 0.05		1	6
ApU	0.36	0.061	< 0.05		4	3
CpG	0.37	0.045	< 0.05		3	4
GpC	0.24	0.041	< 0.05	2	5	
	% (+,0,-)			0.27	0.41	0.32
	Sum - 112		Sum	30	46	36

Legend: Table 12

The sequences of several control regions in _E_. _coli_ and coliphage DNA have been analyzed. The number of pairs of four different types* of dinucleotide were counted in each region. The frequencies of the dinucleotides were calculated as the fraction of the number of bases sequenced in the region.

a. _Lac_ operator region in _E_. _coli_.
   Lac operator-promoter region in _E_. _coli_.
   Gilbert and Maxam, 1973

b. Prom-oter and terminator region of _E_. _coli_
   Tyr-tRNA gene.
   Sekiya and Khorana, 1974; Loewen _et_ _al_., 1974

c. P_L promoter region in lambda.
   Maniatis _et_ _al_., 1974

d. Phage fd binding site for RNA polymerase
   and initiation site for transcription.
   Sugimoto _et_ _al_., 1975

e. A_3 promoter region in phage T7.
   Pribnow, 1975

---

* The dinucleotides were divided into four groups, as follows:

Purines: ApA, ApG, GpA, GpG
Pyrimidines: TpT, TpC, CpT, CpC
Mixed, complementary: CpA, TpG, GpT, ApC
Mixed, self-complementary: TpA, ApT, CpG, GpC

Sequences of regions analyzed

* mRNA synthesis
5'-AATCATGGTCAGAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATT-3' a.
* repressor ginding *

5'-AGTTCAGGGACTTTTGAAAGTGA-3' Terminator region b.

5'-CGCCGCGCAGTAAACTATACTACGCGGGG-3' Promoter region

5'-GTGCTCAGTATCACCGCCAGTGGTATTTATGTC-3' c.

5'-AAATCAAAAATCAGGTCTTTACCCTGTCTATTATAGTCAGAAGC-3' d.
mRNA initiation *

5'-TGACTCACTGTCGTTTCATGTGGTACACGTACCGTGTTTACTT-3' e.
mRNA initiation *

Table 12: Calculated dinucleotide frequencies in control sequences of E. coli and coliphage DNA

	Region Sequenced		
Dinucleotide	[a]Operator *lac*	[a]Operator-Promoter *lac*	[b]Promoter *E. coli* Tyr-tRNA
Purine			
# of pairs	2	5	6
frequency	0.10	0.12	0.21
Pyrimidine			
# of pairs	7	13	3
frequencyq	0.35	0.32	0.11
Mixed, complementary			
# of pairs	5	13	5
frequency	0.25	0.32	0.18
Mixed, self-complementary			
# of pairs	6	9	14
frequency	0.30	0.22	0.50
# of bases sequenced	24	41	29

Dinucleotide frequency in *E.coli*	
Purine	0.237
Pyrimidine	0.243
Mixed, complementary	0.251
Mixed, self-complementary	0.269

Region sequenced				
[b]Terminator E. coli Tyr-tRNA	[c]Lambda P_L Promoter	[d]fd binding site for RNA polym.	[d]fd initiation site	[e]Promoter A_3 in T7
10 0.45	2 0.06	13 0.30	8 0.38	2 0.05
6 0.27	8 0.25	13 0.30	6 0.29	13 0.30
6 0.27	12 0.38	9 0.21	6 0.29	21 0.49
0 0.0	10 0.31	8 0.19	1 0.05	7 0.16
23	32	44	22	44

Dinucleotide frequency in fd plus DNA		
Purine	0.215	(ApA,ApG,GpA,GpG)
Pyrimidine	0.318	(TpT, TpC,CpT,CpC)
Mixed, complementary	0.221	(CpA,TpG,ApC,GpT)
Mixed, self-complementary	0.250	(TpA,ApT,CpG,GpC)

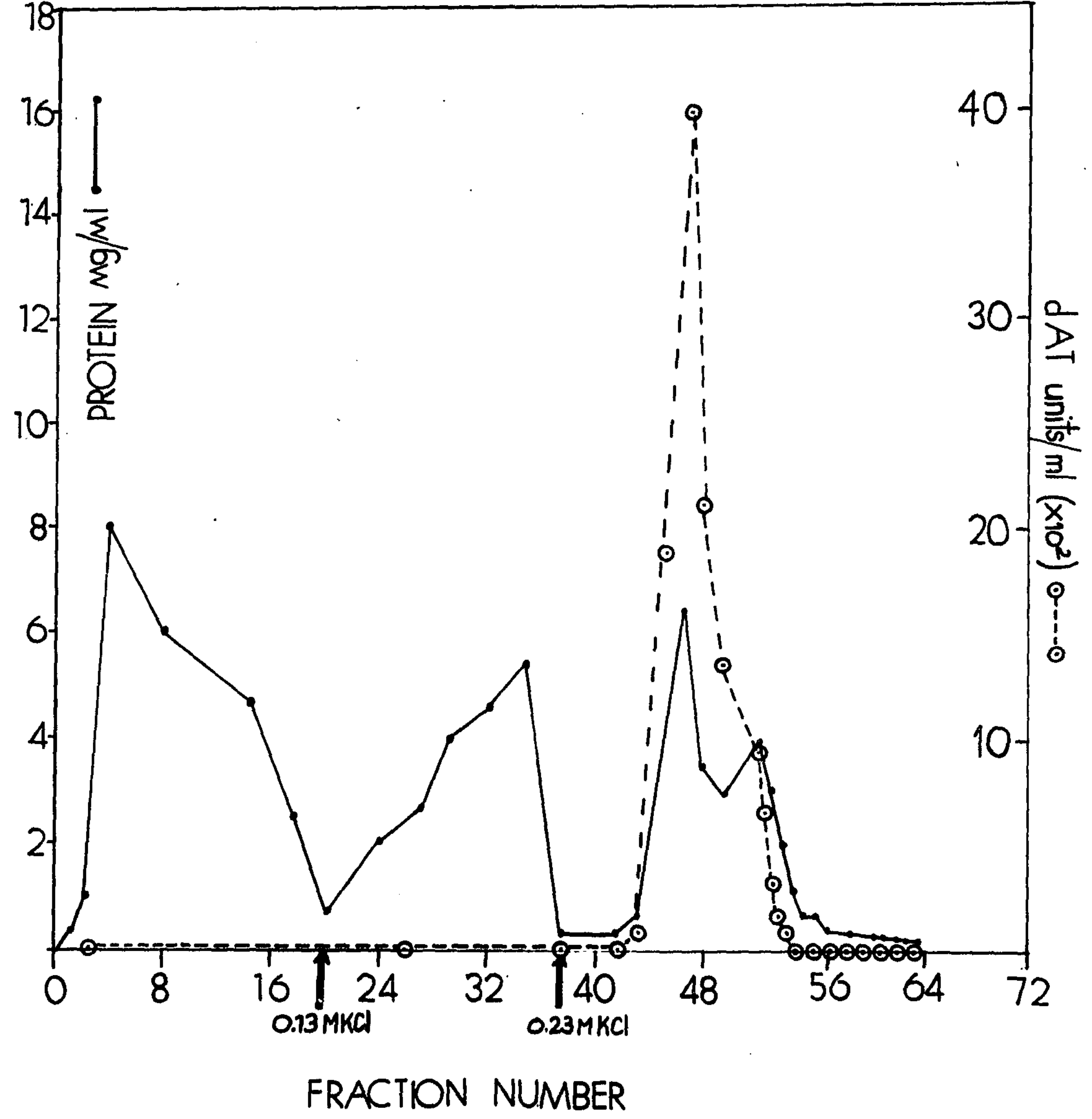
Legend: Figure 1.

Fractionation of RNA polymerase on a DEAE-cellulose column

1780 mg of Fraction III protein was applied to a DEAE-cellulose column (3.5 x 11 cm) that had been equilibrated with buffer A. The protein was eluted in a stepwise manner, as described in the text. The flow rate was 60 ml/hr, and the RNA poltmerase activity was eluted at the 0.23 M KCl step. Protein concentration was determined by the method of Lowry et al. (1951). RNA polymerase activity in each fraction was determined in the standard enzyme reaction (Materials and Methods II), using 4.5 ug of dAT copolymer as template, and ^{14}C-ATP as the labeled substrate (910 cpm/nmole).

Fractions #1-17 contained the flow-through material. Fractions #24-36 contained the buffer A + 0.13 M KCl peak. The RNA polymerase activity was eluted in the buffer A + 0.23 M KCl peak, fractions #46-54. Fractions 46-49, containing the peak enzyme activity were pooled.

FIGURE 1


18
16
14
12
10
8
6
4
2
PROTEIN mg/ml
40
30
20
10
dAT units/ml (x10²)
0
8
16
24
32
40
48
56
64
72
0.13 M KCl
0.23 M KCl
FRACTION NUMBER

Legend: Figure 2.

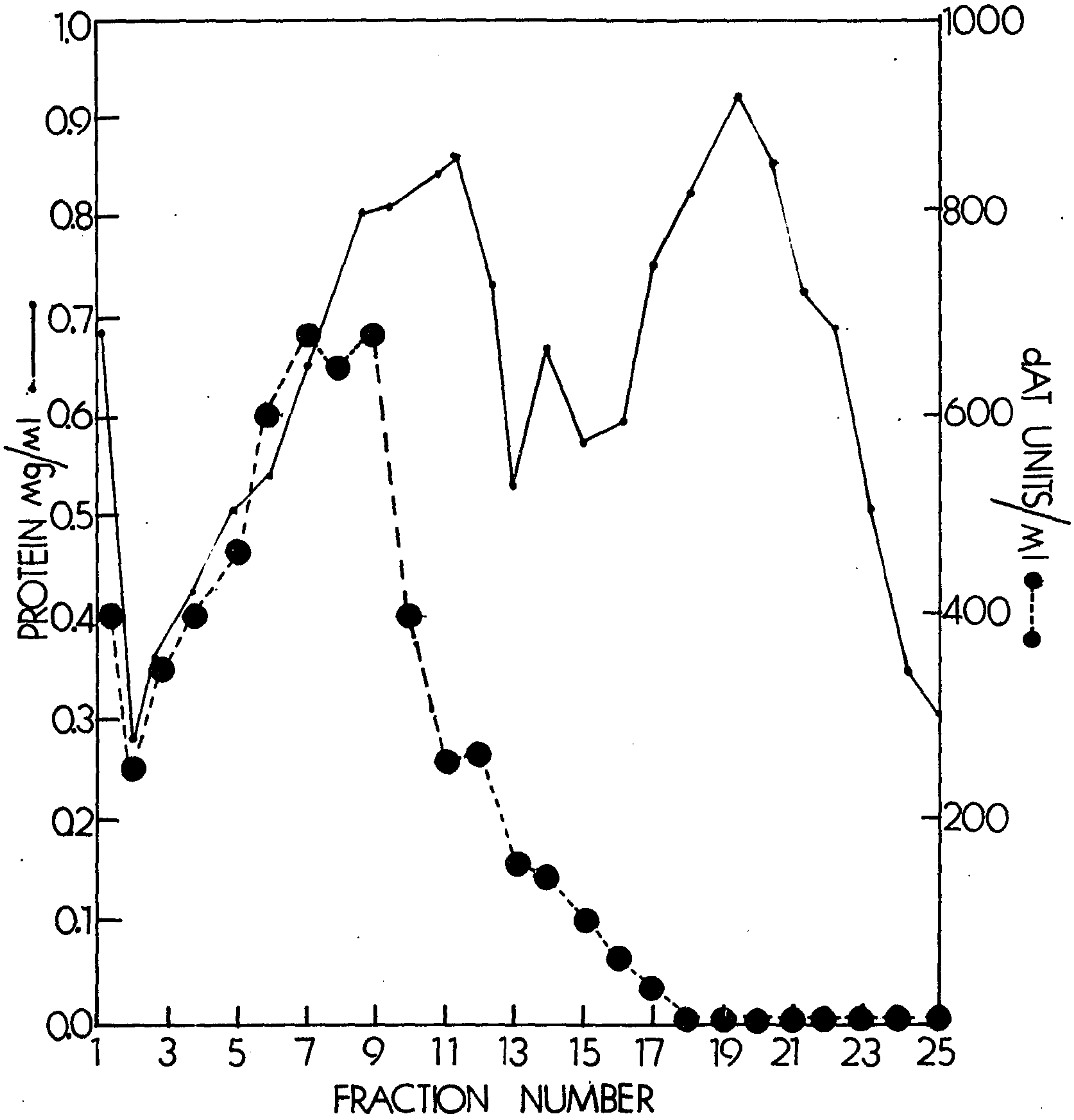
Low salt glycerol density gradient of RNA polymerase

Twenty-five mg of DEAE-cellulose pooled protein (in 1.5 ml) was layered on a 29 ml, 10%-30% glycerol density gradient. The gradient was centrifuged in a Spinco SW 25.1 rotor at 25,000 rpm for 23 hours at 4 C. 1.2 ml fractions were collected.

RNA polymerase activity was determined in the standard enzyme reaction (Materials and Methods II), using 15 nmoles of dAT copolymer as template, and ^{14}C-ATP (910 cpm/nmole) as the labeled substrate. Protein concentration was determined by the method of Lowry _et al_. (1951).

Fractions #16-25, containing the polymerase activity, were pooled.

FIGURE 2

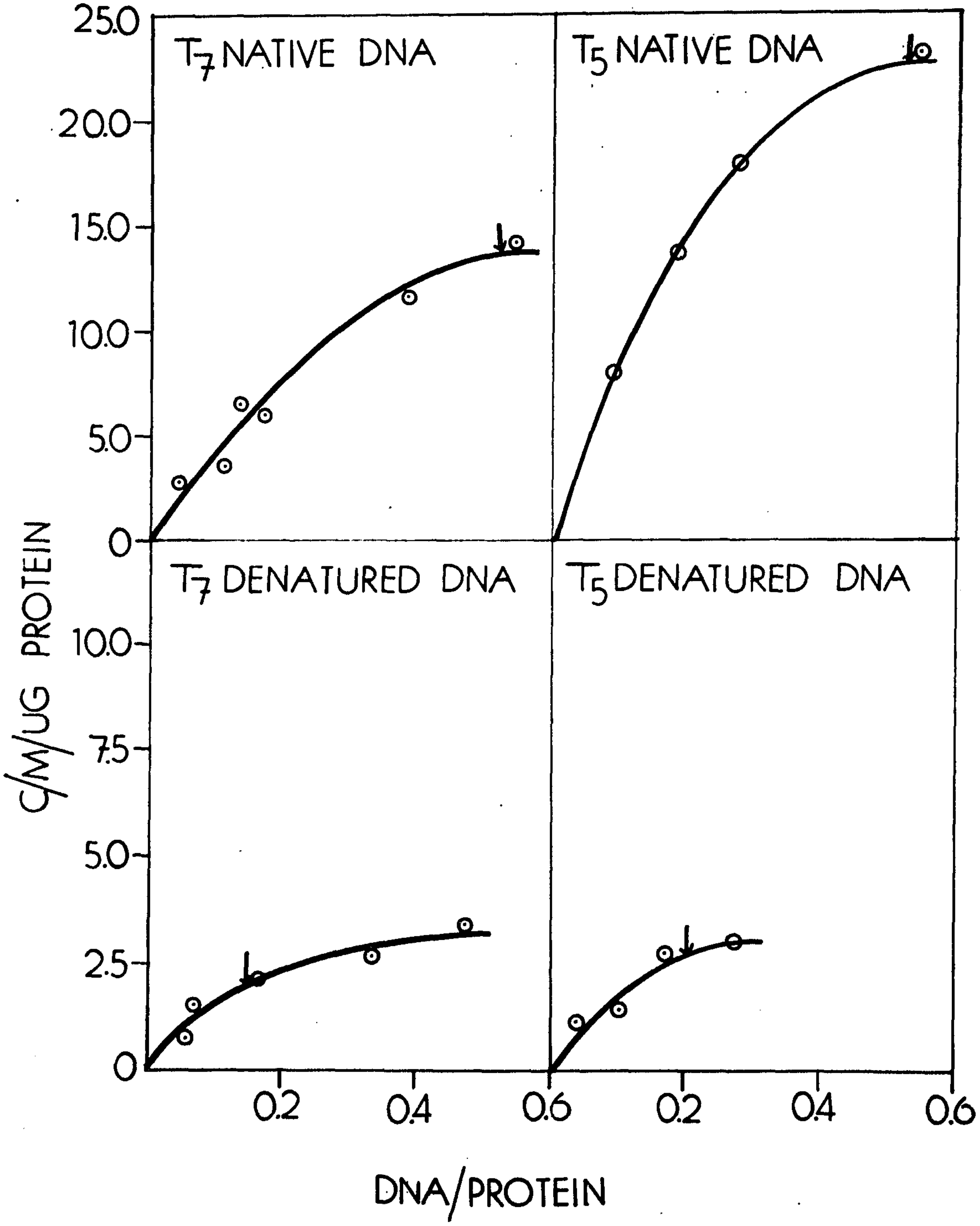

1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
PROTEIN Mg/Ml
1000
800
600
400
200
dAT UNITS/Ml
1
3
5
7
9
11
13
15
17
19
21
23
25
FRACTION NUMBER

Legend: Figure 3.

Saturation curves for native and denatured T5 and T7 phage DNA

Saturation curves were made, using the standard enzyme reaction (Materials and Methods II), containing from 0.5 to 9.0 units of enzyme, and 1 to 20 nmoles of phage DNA. The arrows indicate the ratio of DNA to protein, selected for transcription of each type of DNA. The curves are discussed in the text, Results I.

FIGURE 3


T7 NATIVE DNA
T5 NATIVE DNA
T7 DENATURED DNA
T5 DENATURED DNA
25.0
20.0
15.0
10.0
5.0
0
10.0
7.5
5.0
2.5
0
0.2
0.4
0.6
0.2
0.4
0.6
C/M/UG PROTEIN
DNA/PROTEIN

Legend: Figures 4-7: Dinucleotide frequencies of *in vitro* RNA transcripts of T4, T5 and T7 phage DNA, compared with the values predicted from random association:

Figure 4: ApA, UpU, GpG, CpC
Figure 5: ApG, CpU, GpA, UpC
Figure 6: CpA, UpG, GpU, ApC
Figure 7: UpA, ApU, CpG, GpC

In each figure, the observed nearest-neighbor frequency for each dinucleotide sequence has been plotted against the value predicted from random association. The values for the transcripts of native and denatured DNA are shown, for each DNA studied. The 45 degree line, on each graph, represents the points for which the observed and random frequencies are the same. The values used are from Tables 5D, 6D, 7D and 8D.

Key: Dinucleotide frequencies of:

▲	Native T7 RNA transcripts
△	Denatured T7 RNA transcripts
●	Native T4 RNA transcripts
⊘	Native T4 RNA transcripts (limited-synthesis)
○	Denatured T4 RNA transcripts
◎	T4 DNA , complementary dinucleotide
■	Native T5 RNA transcripts
□	Denatured T5 RNA transcripts
▣	T5 DNA , complementary dinucleotide

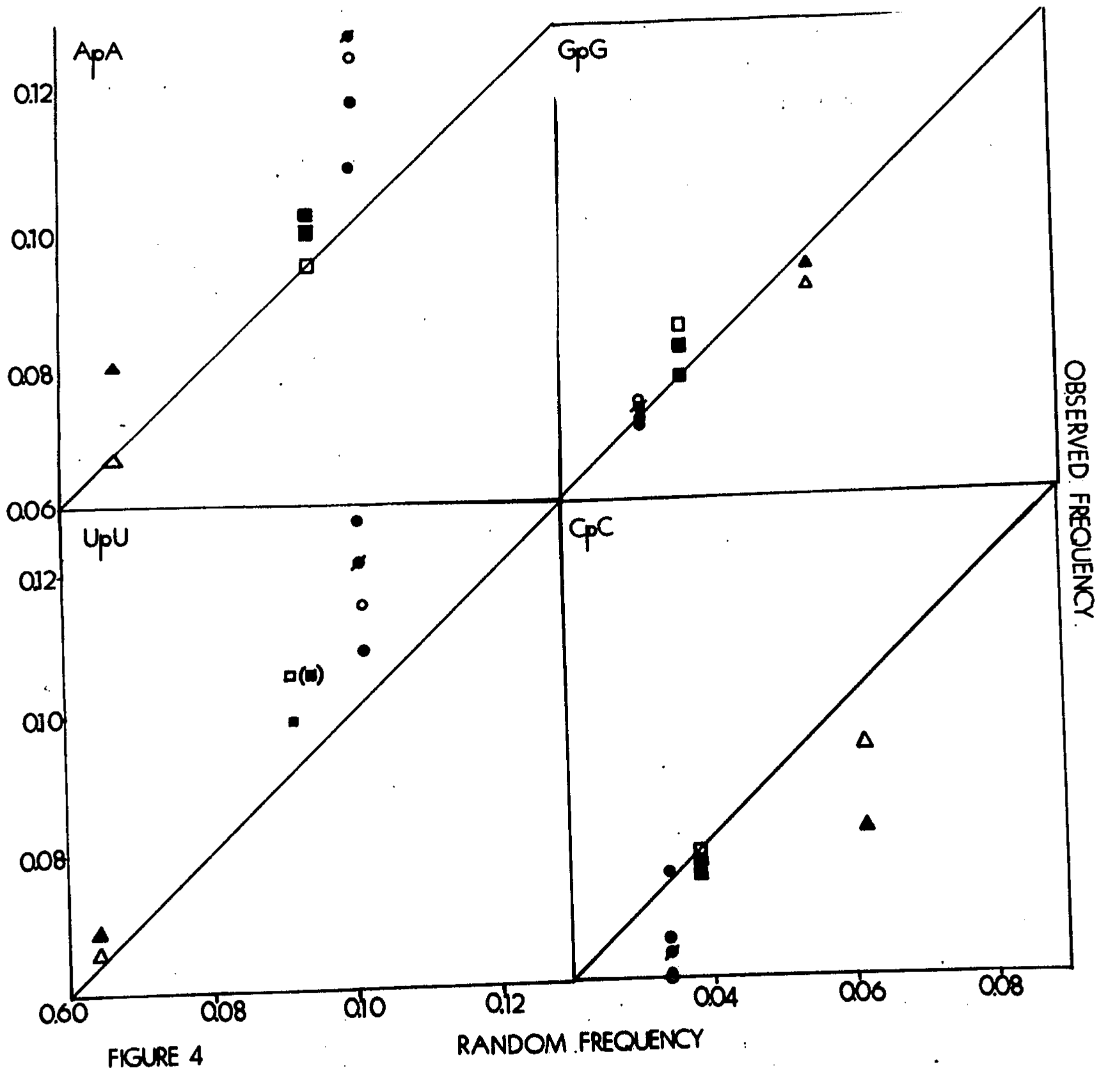

ApA
GpG
UpU
CpC
0.12
0.10
0.08
0.06
0.12
0.10
0.08
0.60
0.08
0.10
0.12
0.04
0.06
0.08
RANDOM FREQUENCY
OBSERVED FREQUENCY

FIGURE 4

FIGURE 5.

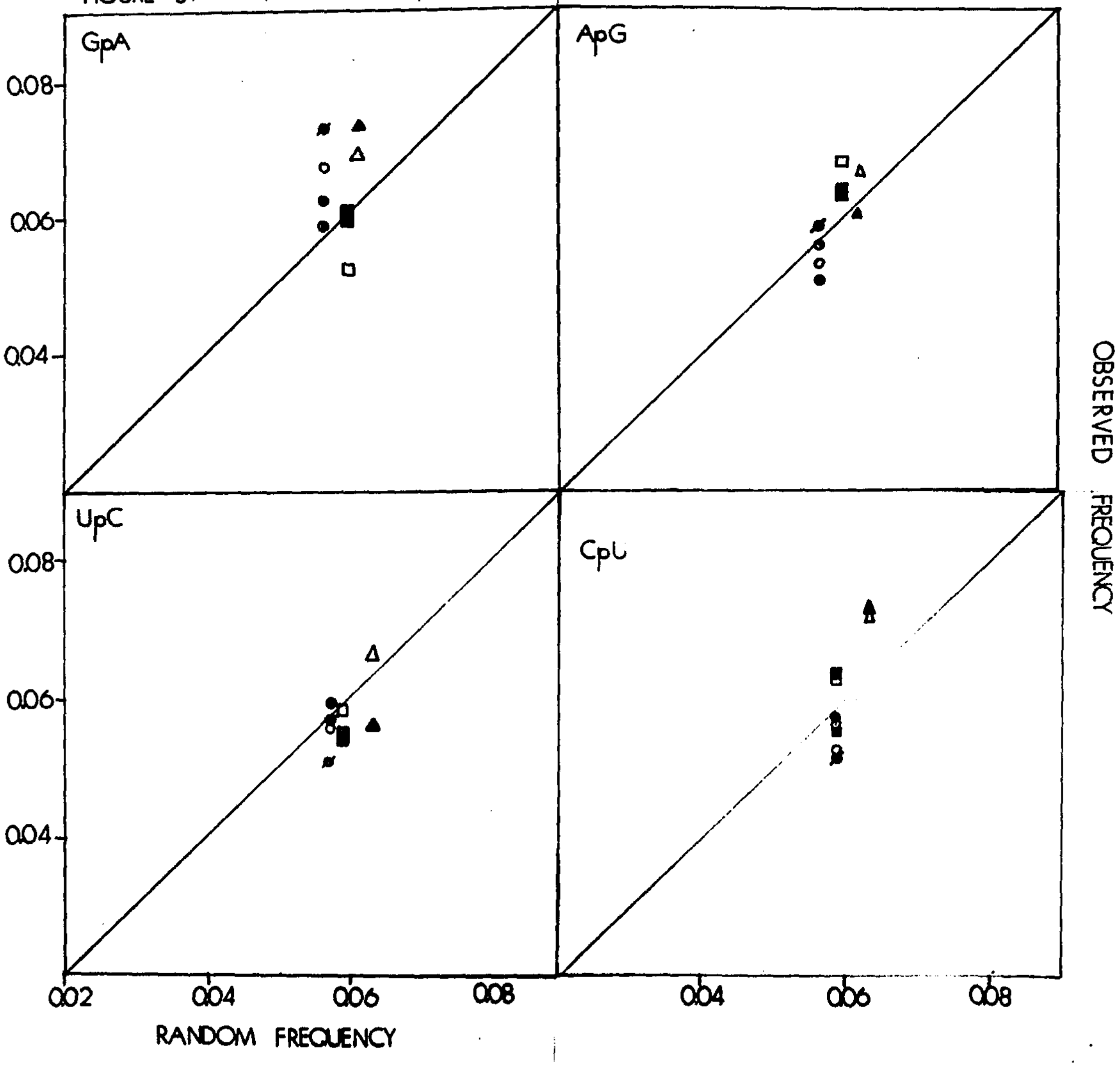

GpA
ApG
UpC
CpU
0.08
0.06
0.04
0.02
0.04
0.06
0.08
RANDOM FREQUENCY
OBSERVED FREQUENCY

FIGURE 6

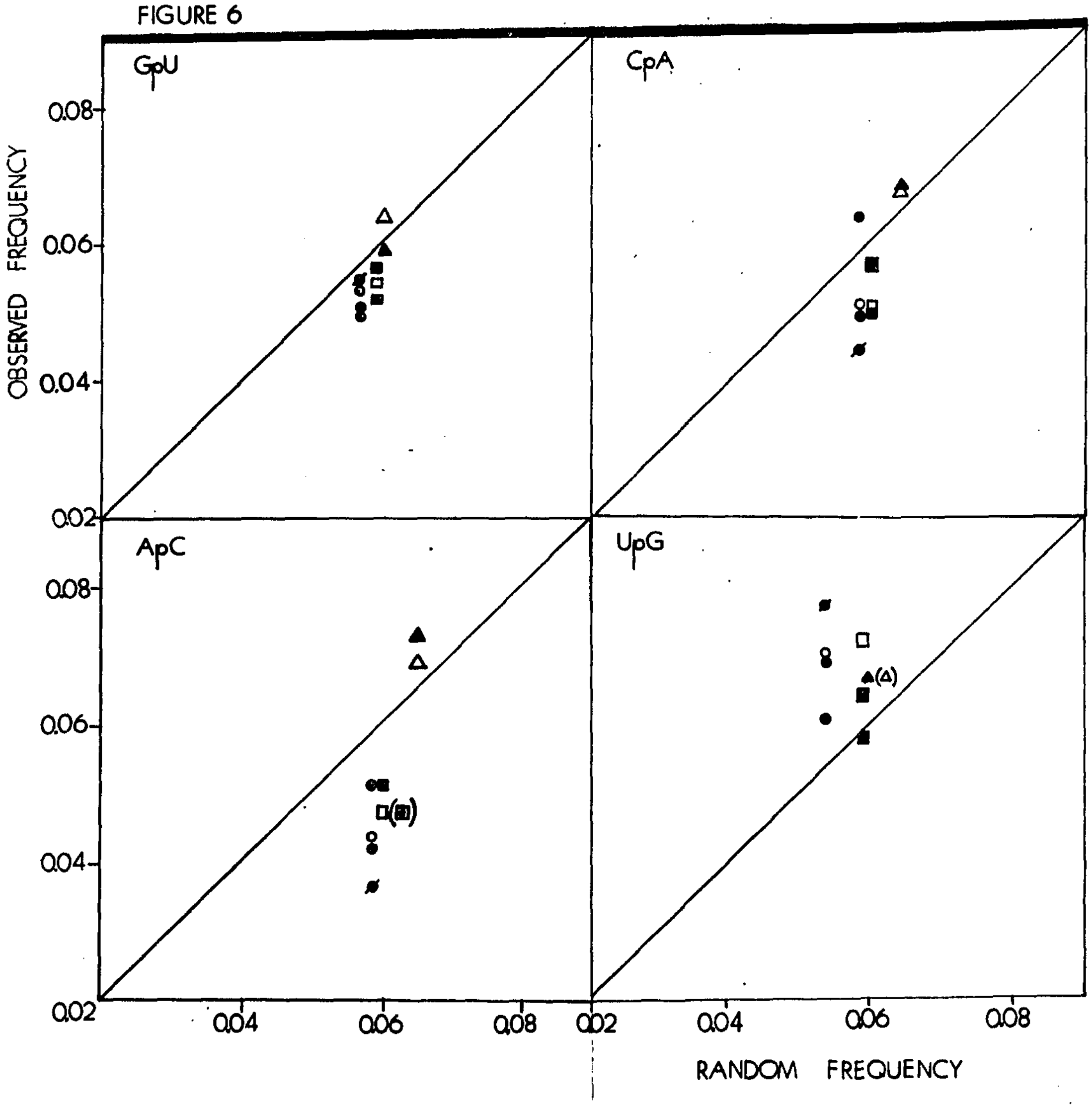
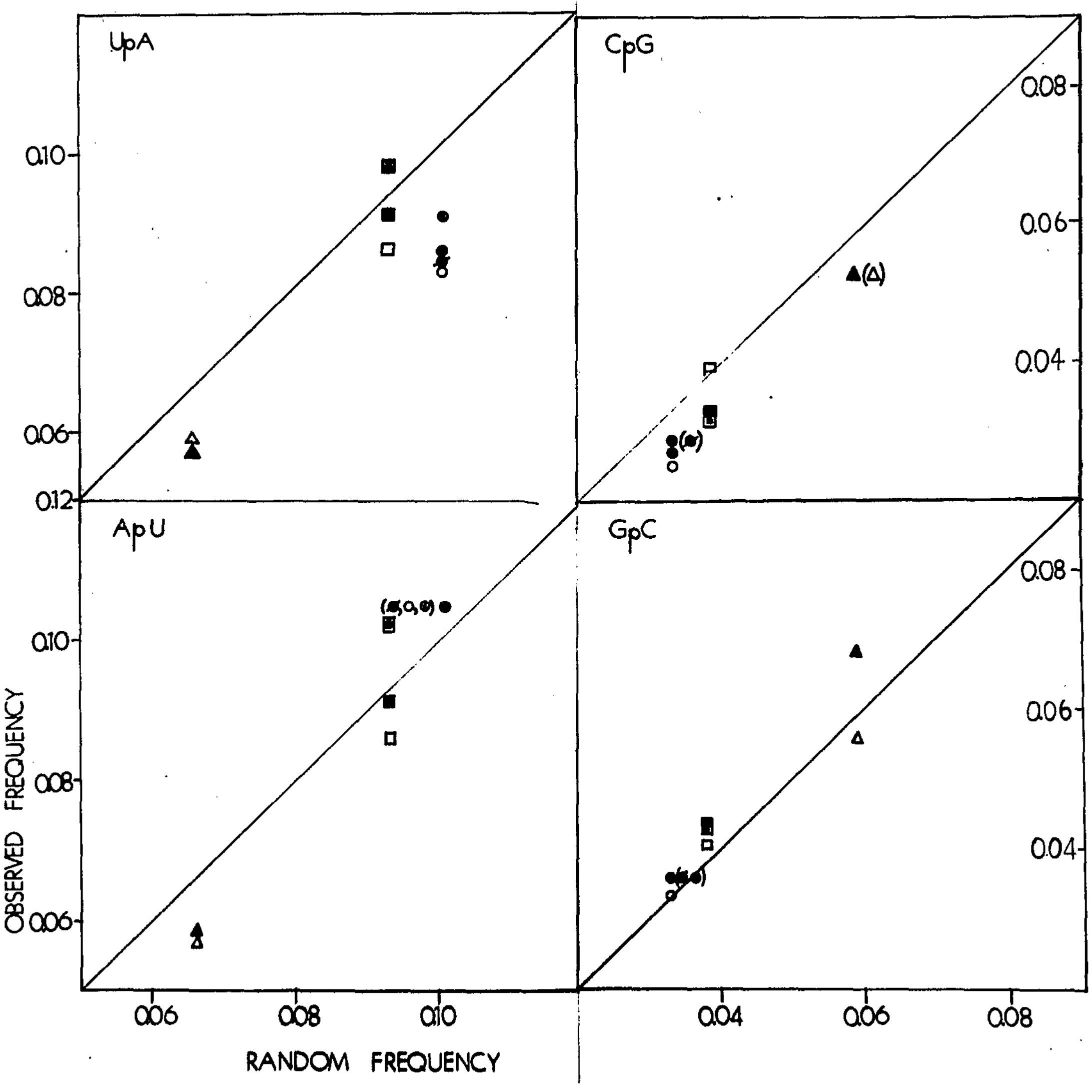

GpU
CpA
ApC
UpG
0.08
0.06
0.04
0.02
0.04
0.06
0.08
(Δ)
OBSERVED FREQUENCY
RANDOM FREQUENCY

FIGURE 7


UpA
CpG
ApU
GpC
0.10
0.08
0.06
0.12
0.08
0.06
0.04
0.10
0.08
0.06
0.08
0.06
0.04
(■,○,●)
▲(△)
(●)
(■,●)
OBSERVED FREQUENCY
0.06
0.08
0.10
0.04
0.06
0.08
RANDOM FREQUENCY