# Nucleotide Clusters in Deoxyribonucleic Acids: Sequence Analysis of DNA Using Pyrimidine Oligonucleotides as Primers in the DNA Polymerase I Repair Reaction†

John S. Kaptein‡ and John H. Spencer*

ABSTRACT: Pyrimidine oligonucleotides have been shown to prime the *E. coli* DNA polymerase I repair reaction, specifically and reproducibly. DNA molecules up to 30 nucleotides long have been obtained from the extension of oligopyrimidine primers, 9 to 11 nucleotides long isolated from the complementary (minus) strand of bacteriophage S13 RF DNA using S13 viral DNA as the template molecule. The sequences of the extended primers were determined from mobility shifts following separation of partially extended primers by ionophoresis and homochromatography, and by a modification of the "plus" system of Sanger and Coulson (1975). The 3′→5′ exonuclease activity of *E. coli* DNA polymerase was utilized for the "plus" system in the presence of single dNTPs and also with two dNTPs in the reaction, to give a nearest neighbor type of analysis for sequence confirmation. The ready availability of oligopyrimidine primers from any DNA and the simplification of the "plus" method broaden the range of applicability of the primed DNA polymerase I repair reaction for DNA sequence analysis.

The repair reaction catalyzed by DNA polymerase I, in which an oligodeoxynucleotide primer bearing a 3′-hydroxyl group is hybridized to a DNA template, then extended with high fidelity of the template sequence, has been demonstrated as a rapid and accurate method of sequence analysis (Sanger and Coulson, 1975). Primers that have been used successfully are restriction fragments and synthetic oligonucleotides (Sanger and Coulson, 1975; Sanger et al., 1977). One type of potential primer which is readily available from any DNA is a pyrimidine oligonucleotide. The long unique pyrimidine oligonucleotide $C_9T_{11}$ present in fl, fd, and M13 DNAs (Petersen and Reeves, 1969; Ling, 1972a; Tate and Petersen, 1974) has been shown to prime DNA polymerase I repair synthesis (Oertel and Schaller, 1972); however, there are conflicting reports on the priming efficiencies of shorter pyrimidine oligonucleotides of chain length 8–12 nucleotide residues (Goulian et al., 1973; Oertel and Schaller, 1973). The priming efficiency of polypyrimidines is low compared with polypurines and the base composition, presumably due to sequence complementarity, affects the efficiency indicating specificity in the priming event (Oertel and Schaller, 1973).

In the present study we have used the replicative form I (RF I)[1] DNA of bacteriophage S13 as the source of pyrimidine oligonucleotide primers, and the viral (plus) strand DNA as the template since oligonucleotides unique to the minus strand of the RF DNA are complementary to unique sites in the viral (plus) strand of the DNA (Cerny et al., 1969). We have shown that pyrimidine oligonucleotides of chain length 9 to 11 nucleotides are effective as specific primers in the DNA polymerase I repair reaction and have demonstrated their use for sequence analysis studies. A preliminary report of this work has appeared elsewhere (Kaptein and Spencer, 1976).

## Materials and Methods

*Chemicals and Enzymes.* The sources and purity of all chemicals used have been described elsewhere (Harbers et al., 1976). Bacterial alkaline phosphatase (electrophoretically pure) and snake venom phosphodiesterase were obtained from Worthington Biochemical Corp. and treated to remove phosphodiesterases (Garen and Levinthal, 1960) and 5′-nucleotidase activity (Sulkowski and Laskowski, 1971), respectively. Polynucleotide kinase was a generous gift of Dr. Allen Delaney and its preparation has been described elsewhere (Harbers et al., 1976).

DNA polymerase I was a generous gift of Dr. Frank Grosveld and was prepared by modifications (Grosveld, 1977) of the Richardson et al. (1964) procedure as follows: *E. coli* B was grown, the cells were harvested and lysed (Richardson et al., 1964), and nucleic acids removed from the extract by liquid-polymer phase partitioning using poly(ethylene glycol) and dextran (Kamen, 1972). The proteins in the top phase were precipitated by addition of solid ammonium sulfate to saturation, over a 30-min period. The ammonium sulfate solution was stirred for an additional 20 min at 0 °C and then centrifuged at low speed (20 min, 10 000 rev/min, IEC rotor no. 872). The precipitated protein which formed a cake at the interphase between the poly(ethylene glycol) and ammonium sulfate solution phases was collected, dissolved in 50 mM potassium phosphate buffer (pH 7.5) containing 3 mM $MgCl_2$, 1 mM EDTA, and 0.1 mM glutathione, and the solution dialyzed overnight against the same buffer except that the glutathione concentration was increased to 0.1 M. The dialysate was clarified by centrifugation (50 min, 10 000 rev/min IEC rotor no. 870) and the supernatant fractionated with ammonium sulfate as described by Richardson et al. (1964). The DNA polymerase I fraction was collected by centrifugation, dissolved in 20 mM potassium phosphate buffer (pH 7.2), and dialyzed overnight against 0.2 M potassium phosphate buffer (pH 6.5), containing 10 mM glutathione. The enzyme-containing solution was then chromatographed on phosphocellu-

† From the Department of Biochemistry, McGill University, Montreal, Quebec, Canada H3G 1Y6. *Received May 18, 1977; revised manuscript received November 7, 1977.* This is paper 15 of a series. Paper 14 is Hewish et al. (1976). This work was supported by Grant MT 1453 from the Medical Research Council of Canada.

‡ Holder of a Medical Research Council of Canada Studentship.

[1] Abbreviations used: RF, replicative form; TEAB, triethylammonium bicarbonate; *Hin*d, *Hae*III, and *Pst*I, restriction enzymes from *Hemophilus influenzae, Hemophilus aegyptius,* and *Providencia stuartii,* respectively.

0006-2960/78/0417-0841$01.00/0 

lose and Sephadex G-100 as described by Jovin et al. (1969). The fractions containing the enzyme were pooled, dialyzed against 0.1 M potassium phosphate buffer (pH 7.0) in 50% glycerol, and stored at −20 °C. The dialyzed enzyme solution assayed according to Richardson et al. (1964) contained 260 units/mL.

*Purification of Unlabeled Bacteriophage S13 Wild Type DNA.* This was performed according to Spencer and Boshkov (1973).

*Purification of Unlabeled and $^{32}$P-Labeled Bacteriophage S13 RF DNA.* Growth was performed as described by Shleser et al. (1968). After collection of the cells by centrifugation, the replicative form DNA was isolated from the total DNA as described by Schekman et al. (1971) except for two modifications. After isopropyl alcohol precipitation, the nucleic acid precipitate was redissolved in 50 mM Tris-HCl buffer, pH 8.1, containing 0.15 M NaCl and 10 mM EDTA and applied to a Bio-Gel A-15 m column. The column was eluted with the same Tris-NaCl-EDTA buffer and the DNA which eluted in the void volume collected and fractionated in an ethidium bromide–CsCl density gradient. The ethidium bromide was removed from the pooled gradient fractions as described by Grosveld et al. (1976).

*Isolation of Pyrimidine Oligonucleotide Primers.* Pyrimidine oligonucleotides were released from bacteriophage S13 RF DNA by treatment of the DNA with 2% diphenylamine in 67% formic acid at 30 °C for 18 h (Burton, 1967). The hydrolysate was extracted five times with ether, dried in a desiccator under vacuum, and dephosphorylated with alkaline phosphatase as described by Delaney and Spencer (1976). The dephosphorylated oligonucleotides were separated according to chain length and base composition by fractionations on DEAE-Sephadex and then desalted (Delaney and Spencer, 1976). The oligonucleotides were stored in 5 mM Tris-HCl buffer (pH 8.0) at −20 °C at a concentration of 300 pmol/mL.

*5′-Terminal Labeling of Pyrimidine Oligonucleotide Primers.* Oligonucleotide solutions containing approximately 5 pmol of oligonucleotide (15 μL) were mixed with 50 μL of 0.1 M Tris-HCl buffer (pH 8.9) containing 20 mM $MgCl_2$, 20 μL of [γ-$^{32}$P]ATP (approximately 10–20 × $10^6$ dpm), 4 μL of β-mercaptoethanol, and 10 μL (0.006 unit) of polynucleotide kinase. The mixture was incubated overnight at room temperature, then the 5′-$^{32}$P-labeled oligonucleotides were separated from the excess [γ-$^{32}$P]ATP by chromatography on Sephadex G-25 packed in a 1-mL disposable pipet, and eluted with 5 mM TEAB, pH 8.0. The oligonucleotides eluted in the void volume which was collected and pooled.

*The DNA Polymerase I Repair Reaction.* To effect extension of the 5′-$^{32}$P-labeled primer in the DNA polymerase I repair reaction so that a complete collection of intermediate extension products, each differing by one nucleotide, was present in the final reaction product, a series of limiting reaction conditions was devised. The rationale was that a limiting concentration of one triphosphate would slow down the primer extension polymerization reaction. Sequential aliquots could then be taken at selected times providing a spectrum of partial extension products. A limiting reaction was set up with each of the four triphosphates to circumvent the possibility of sequence effects and variations on the extension reaction.

The solution of terminally 5′-$^{32}$P-labeled pyrimidine oligonucleotide primer eluted from Sephadex G-25 was divided equally into four microtubes. To each tube was added 10 μL of a solution containing 65 mM Tris-HCl buffer (pH 7.0), 4 mM $MgCl_2$, 1 mM dithiothreitol, and 50 mM NaCl, then 1 μL of a solution of 1 mM of each of three appropriate deoxynucleoside triphosphates and 1 μL of a 10 μM solution of the fourth deoxynucleoside triphosphate. A different deoxynucleoside triphosphate of the 10 μM concentration was used in each of the four sample tubes. Upon completion of the additions, the solutions were dried under vacuum, the dried samples each dissolved in 5 μL of 5 mM Tris-HCl buffer (pH 7.5) containing 2.5–3 mg/mL of S13 viral DNA, and the mixtures heated in a boiling water bath for 3 min to disrupt any possible secondary structure. After preincubation at room temperature overnight to promote hybridization of the oligonucleotide primer to the viral DNA template (Oertel and Schaller, 1973), 3–5 μL (0.75–1.25 units) of a solution of DNA polymerase I was added to each tube and the mixtures were incubated at room temperature for 2 h. During the incubation sequential, samples of 1 μL were removed from each tube starting at 30 min after the onset of incubation and continuing at 20-min intervals to the maximum 2 h. As they were removed, all the 1-μL samples were pooled by addition to a 4-μL solution of 0.5 M EDTA held at 0 °C. The pooled samples were deproteinized by phenol extraction. Phenol was removed by extraction three times with ether and the elongated primer solution applied to Sephadex G-25 packed in a 1-mL disposable pipet, to remove the unused deoxynucleoside triphosphates. The column was eluted with 5 mM TEAB buffer (pH 8.0) and the extended primer which was in the void volume collected.

*Fractionation and Sequence Analysis of Extended Primer Products.* Products of the extension reaction were fractionated either by ionophoresis on cellulose acetate or Cellogel followed by homochromatography on DEAE-cellulose thin-layer plates (Brownlee and Sanger, 1969), or unidimensionally by homochromatography only. Thin-layer plates were coated with a mixture of DEAE-cellulose and cellulose in a ratio of 1:7.5 and elution was with a 2% solution of partially hydrolyzed yeast RNA at 60 °C (Jay et al., 1974). The separated oligonucleotides were visualized by autoradiography on Kodak BB14 x-ray film.

For sequence analysis the oligonucleotide spots were scraped loose from the TLC plates and collected by vacuum into Pasteur pipet tips plugged with cotton. The scrapings were washed with cold ethanol to remove urea and the oligonucleotides eluted by 1.3 M TEAB (pH 8.0) at 90 °C. The eluate was dried under vacuum and redissolved in water several times to remove traces of TEAB. The oligonucleotides were partially digested with snake venom phosphodiesterase (Ling, 1972b; Harbers et al., 1976) and the products separated by ionophoresis–homochromatography (Brownlee and Sanger, 1969).

*The DNA Polymerase I Repair Reaction in a Two-Stage Extension Procedure.* The theoretical basis of this experiment is that the extended primer is degraded by the 3′→5′ exonuclease activity of the DNA polymerase I, until the nucleotide(s) which is present in the reaction mixture is reached in the sequence. At this point the polymerization activity of the enzyme supercedes the exonucleolytic activity and prevents further degradation from the 3′ end resulting in an accumulation of primer extension products at this point in the sequence.

The 5′-$^{32}$P-labeled oligonucleotide primer was first extended by the DNA polymerase I repair reaction as described above. The solution eluted in the void volume from the Sephadex G-25 column was divided into ten equal portions, and to each was added 10 μL of 65 mM Tris-HCl buffer, pH 7.0, containing 4 mM $MgCl_2$, 1 mM dithiothreitol, and 50 mM NaCl. Additions of deoxynucleoside triphosphates were then made according to Table I. All ten samples were dried under vacuum, 5 μL of S13 viral DNA was added, the mixture was annealed as described above, DNA polymerase I was added, and the solutions were incubated at room temperature for 2 h. The

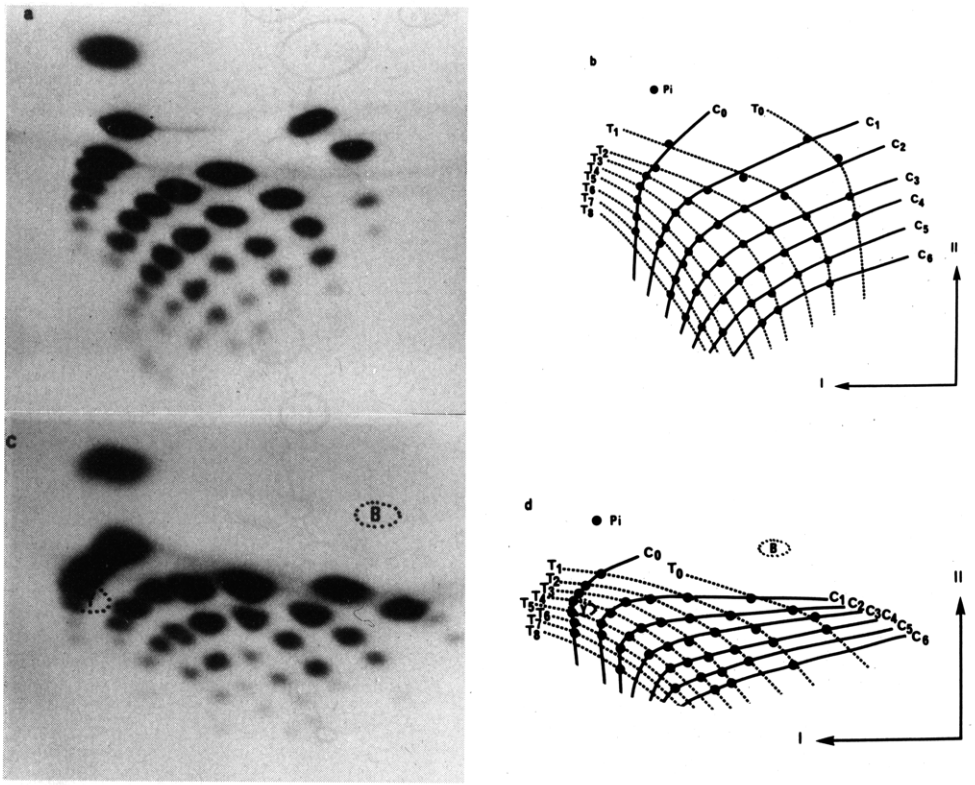


FIGURE 1: Two-dimensional fractionations of the pyrimidine oligonucleotides of bacteriophage S13 viral and RF DNAs. (Dimension I) Ionophoresis at pH 3.5. (Dimension II) Homochromatography on DEAE-cellulose thin-layer plates at 60 °C, eluent 2% partially hydrolyzed yeast RNA containing 7 M urea. The separated oligonucleotides were visualized by autoradiography. (Y) Position of yellow dye marker (orange G); (B) position of blue dye marker (xylene cyanol FF); $P_i$, inorganic phosphate. (a) Pyrimidine oligonucleotides released from uniformly $^{32}P$-labeled S13 RF DNA by formic acid–diphenylamine hydrolysis. (b) Grid pattern derived from Figure 1a. (c) Pyrimidine oligonucleotides released from uniformly $^{32}P$-labeled S13 viral DNA by formic acid diphenylamine hydrolysis. (d) Grid pattern derived from Figure 1c.

TABLE I: Composition of Reaction Mixtures for the Second Stage Extension.

| Component | Volume (μL) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| dATP (1 mM) | 1 | | | | 1 | 1 | 1 | | | |
| dGTP (1 mM) | | 1 | | | 1 | | | 1 | 1 | |
| dCTP (1 mM) | | | 1 | | | 1 | | 1 | | 1 |
| dTTP (1 mM) | | | | 1 | | | 1 | | 1 | 1 |

reactions were stopped by addition of phenol as described above and the oligonucleotide products chromatographed unidimensionally on 20 × 40 cm DEAE-cellulose TLC plates eluted by 2% partially hydrolyzed yeast RNA at 60 °C. The separated oligonucleotides were visualized by autoradiography.

## Results

The autoradiographs in Figures 1a and 1c show the separation by ionophoresis–homochromatography of $^{32}P$-labeled pyrimidine oligonucleotides present in formic acid–diphenylamine hydrolysates of S13 RF DNA and S13 viral (plus) strand DNA, respectively. The graticules in Figures 1b and 1d are the interpretation of the autoradiographs. Comparison of Figures 1a and 1c reveals that oligonucleotide tracts $C_5T$, $C_3T_6$, $C_2T_7$, $C_4T_7$, and $C_3T_8$ are unique to the minus strand and the relative intensity of the spots gives an indication of the relative frequency of occurrence of these oligonucleotide tracts in S13 DNA. These data confirm the previous pyrimidine tract composition and distribution analysis obtained by column fractionation procedures (Cerny et al., 1969; Delaney and Spencer, 1976) and provided the basis for the choice of oligonucleotides to be sequenced and used as primers in the DNA polymerase I repair reaction.

The sequences of the longer pyrimidine oligonucleotides occurring in 1 or 2 molar yield in the minus strand of S13 RF DNA (Figure 1a) were deduced from mobility shifts of products of their partial digestion with either snake venom phosphodiesterase or spleen phosphodiesterase when fractionated by ionophoresis–homochromatography (Ling, 1972b). A compilation of the results is presented in Table II.

Figures 2a, 2b, and 2c are autoradiographs of the separation by ionophoresis–homochromatography of the products of partial snake venom phosphodiesterase digestion of $CT_6$, $C_3T_8$, and $C_4T_7$, respectively. In Figure 2a the products derived from the $CT_6$ oligonucleotide isomer present in the plus strand DNA (Harbers et al., 1976) are joined by a dotted line. The products joined by solid lines are from the single $CT_6$ oligonucleotides present in the minus strand of the RF DNA, and its sequence is T-T-T-T-T-T-C. Oligonucleotide $C_3T_8$ is unique to the minus strand of S13 RF DNA (Figure 1a) but is present in two isomeric copies. The sequential removal of nucleotides from the 3′ end reveals (Figure 2b) that the two sequences are identical for seven nucleotides. Beyond this the sequences diverge and no other digestion product is observed to be common to both sequences. Thus the sequences were determined unambiguously as C-C-T-T-T-C-T-T-T-T-T and T-C-T-C-T-C-T-T-T-T-T. Oligonucleotide $C_4T_7$ is unique to the minus strand of the RF DNA and is present in one copy. From the autoradiograph in Figure 2c the partial sequence is -C-T-T-T-C-C-T-T-T. 5′-terminal nucleotide analysis by labeling with [$\gamma$-$^{32}P$]ATP and polynucleotide kinase followed by total snake venom phosphodiesterase digestion showed the 5′ nucleotide was T (results not shown). The sequence T-C-C-T-T-T-C-

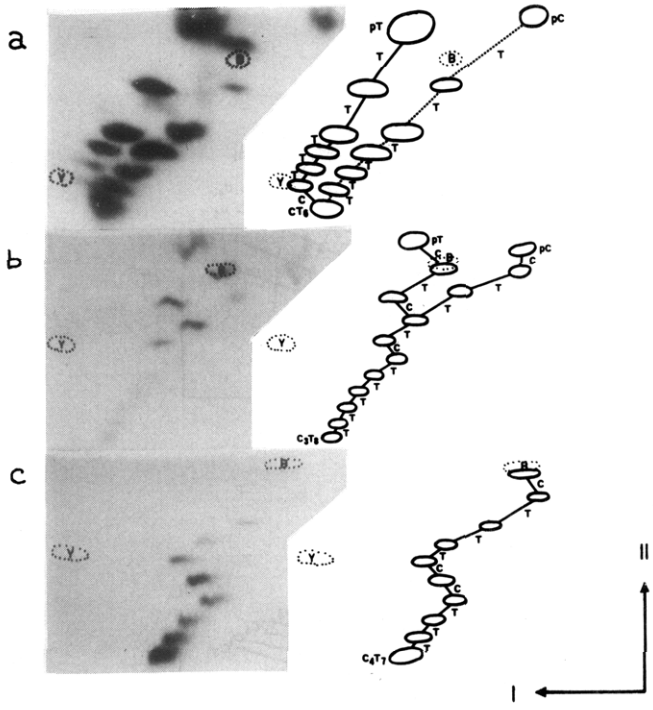


FIGURE 2: Two-dimensional fractionations of the products resulting from partial snake venom phosphodiesterase digestion of 5′-$^{32}$P-labeled pyrimidine oligonucleotides from S13 RF DNA. (a) Oligonucleotide $CT_6$; (b) oligonucleotide $C_3T_8$; (c) oligonucleotide $C_4T_7$. For details of the separation, see legend to Figure 1. The autoradiographs have been schematized for interpretive clarity.

TABLE II: Sequences of Long Pyrimidine Oligonucleotides from the Minus Strand of S13 RF DNA.

| Isostich | Composition | No. of tracts per minus strand molecule | Sequence |
|---|---|---|---|
| 7 | $C_3T_4$ | 1 | Not done[a] |
| | $C_2T_5$ | 2 | Not done[a] |
| | $CT_6$ | 1 | T-T-T-T-T-T-C |
| 8 | $C_5T_3$ | 1 | T-C-C-C-C-T-T-C |
| | $C_4T_4$ | 3 | Not done |
| | $C_2T_6$ | 1 | C-T-C-T-T-T-T-T |
| 9 | $C_5T_4$ | 1 | C-T-T-C-T-C-C-T-C |
| | $C_3T_6$ | 1 | T-C-T-C-T-T-T-C-T |
| | $C_2T_7$ | 2 | T-C-T-C-T-T-T-T-T<br>T-T-T-T-C-T-T-C-T |
| 11 | $C_4T_7$ | 1 | T-C-C-T-T-T-C-C-T-T-T |
| | $C_3T_8$ | 2 | C-C-T-T-T-C-T-T-T-T-T<br>T-C-T-C-T-C-T-T-T-T-T |

[a] The presence of multiple composition isomers derived from the plus strand prevented sequence analysis of these oligonucleotides. Symmetrical sequences six nucleotides and longer, which are true palindromes, are underlined and the center of symmetry indicated by a dot.

C-T-T-T was confirmed by partial spleen phosphodiesterase digestion.

Several of the sequences presented in Table II were confirmed by partial snake venom phosphodiesterase digestion of the products of the initial partial digests as described by

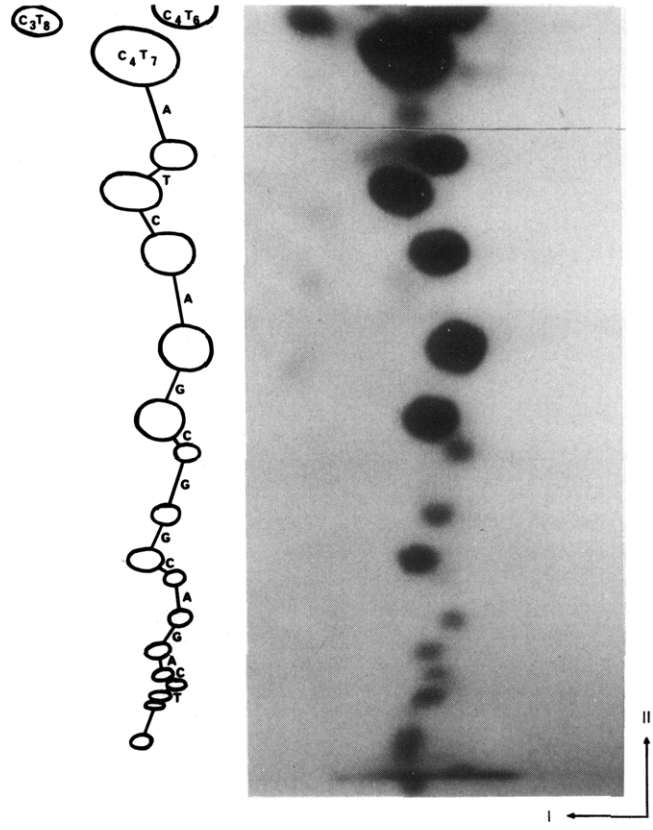


FIGURE 3: Two-dimensional fractionation of the 5′-$^{32}$P-labeled products resulting from extension of oligonucleotide primer $C_4T_7$ in the *E. coli* DNA polymerase I repair reaction, on an S13 viral DNA template. For details of the separation, see legend to Figure 1.

TABLE III: Sequence of Some Extension Products of $C_4T_7$.

| Oligonucleotide (from Figure 4) | Sequence |
|---|---|
| | 5′ 3′ |
| $C_4T_7$ | T-C-C-T-T-T-C-C-T-T-T |
| a | . . . . . T-T-T-C-C-T-T-T-A |
| b | . . . . . T-T-A-T |
| c | . . . . . T-C-A |
| d | . . . . . C-A-G |
| e | . . . . . C-G-G |

Harbers et al. (1976). A large proportion of the sequences revealed the presence of true palindromes, a feature noted previously (Harbers et al., 1976) for the viral strand sequences.

Extension of oligonucleotide $C_4T_7$ when used as a primer in the DNA polymerase I repair reaction is shown in Figure 3. The extension was more than 16 nucleotides and the sequence was determined on the basis of mobility shifts in the ionophoresis–homochromatography system. The presence of $C_3T_8$ on the chromatogram was due to a minor contamination of the $C_4T_7$ starting material, the result of a very slight overlap of the two oligonucleotides when eluted from the DEAE-cellulose column during isolation (Cerny et al., 1969). The $C_4T_6$ spot is due to degradation of the $C_4T_7$ primer by the 3′→5′ exonuclease activity of the DNA polymerase I (see later). Confirmation of the sequence was by elution of selected extension products which were further analyzed by partial snake venom phosphodiesterase digestion; autoradiographs of some of these results are shown in Figure 4. The schematic on the left of Figure 4 indicates the extension products of the $C_4T_7$

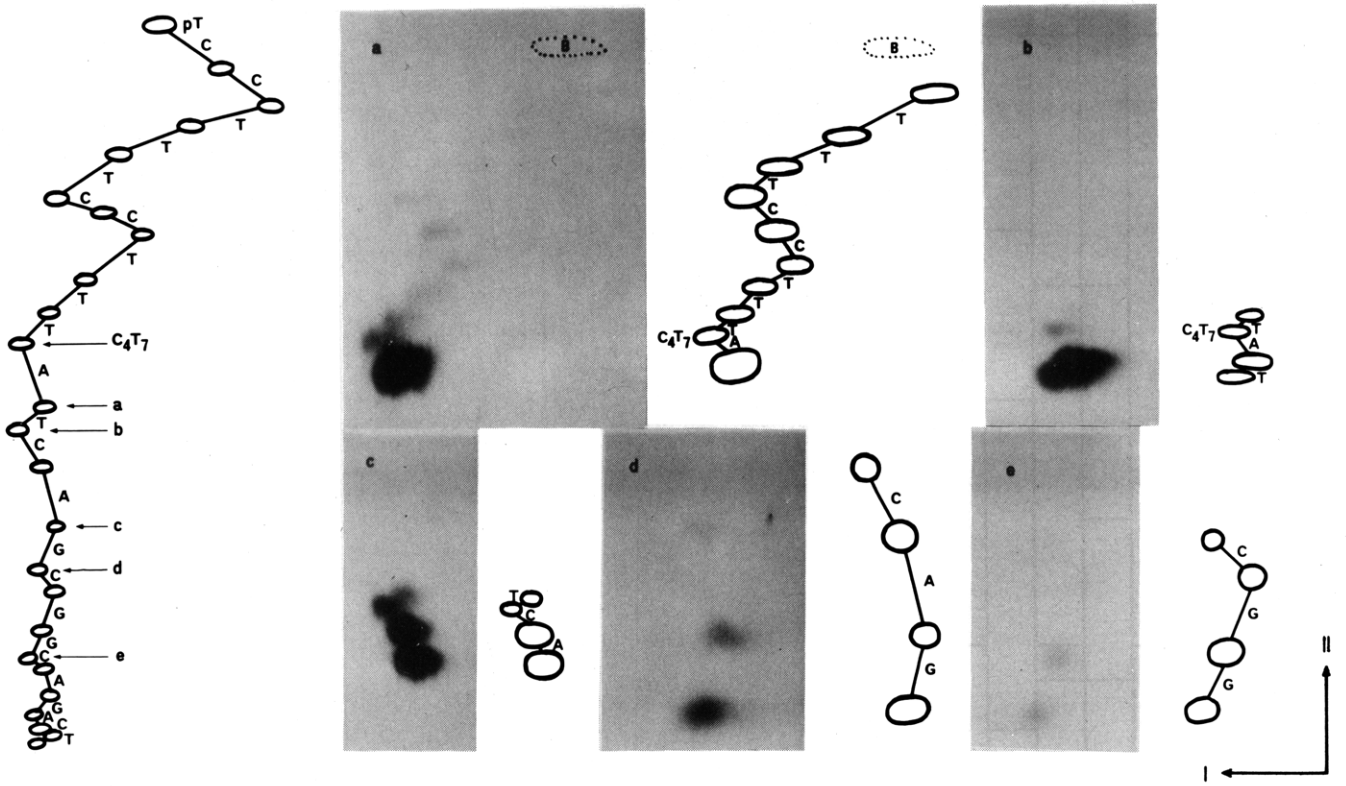


FIGURE 4: Two-dimensional fractionations of the products resulting from partial snake venom phosphodiesterase digestion of selected extension products of primer $C_4T_7$. The schematic of the extension of $C_4T_7$ from Figure 3 is on the left and those extension products hydrolyzed in this experiment are marked a to e inclusive. For details of the separation, see legend to Figure 1.

primer, the products of the action of the 3′→5′ exonuclease component of DNA polymerase on the primer, and those extension products a to e inclusive, which were redigested. The sequences from products a to e are listed in Table III. In all cases except e a sequence overlap with the previous oligonucleotide was obtained.

The redigestion of the primer extension products was also used to confirm that no intermediate extension product was absent from or unobserved on the original autoradiograph of the extension products (Figure 3).

Further confirmation of the sequence was by the two-stage extension of the primer. The results for $C_4T_7$ are presented in Figure 5. The center column of the autoradiograph bordered by vertical lines is a control experiment of the primer extension products after the first extension (see Figure 3).

Interpretation of the autoradiograph is given in the schematics of Figure 5. When dATP is present accumulation of oligonucleotide products occurs at extension positions one and four. When dCTP is present accumulation is at positions three and six, the position of these residues in the extended sequence. When dGTP is present accumulation is at positions one, two, five, and eight. The spots at positions one and two are less intense than those from other extensions (see dATP and dTTP) and are concluded to be anomalous, particularly in view of the first extension data. No accumulation is evident at position seven since this is next to the same base at position eight and this would be blocked from action of the 3′→5′ exonuclease. Similar anomalies have been observed by Sanger and Coulson (1976) with the plus–minus method. When dTTP is present accumulation is at positions one and two. Again the spot at position one is anomalous (see dATP extension and first extension data).

When pairs of nucleoside triphosphates are present, the result is more complex. The first nucleotide addition product accumulates when dATP + dCTP or dATP + dGTP are present. When dATP + dTTP or dGTP + dTTP are present, the second product accumulates, not the first, confirming adenylic acid as the first addition and thymidylic acid as the second. Note that the T in the second position could be predicted since it was the missing triphosphate out of the four possibilities in the two dNTP pairs involved in accumulation of the first addition product. Thus C should be in the third position since it is the missing dNTP in the dATP + dTTP and dGTP + dTTP pairs involved in accumulation of the second addition product. This confirmatory type of evidence can be compared with a nearest neighbor analysis. The confirmed sequence is $C_4T_7$-A-T-C-A-G-C-G-G-C-A-G. Several minor spots are present in the reaction mixtures and are probably due to an incomplete reaction at the second step of the procedure.

Some spots on the chromatogram (Figure 5) are anomalous (spot X) and may be the result of the minor contamination of primer $C_4T_7$ with $C_3T_8$. It is unlikely that $C_4T_7$ is priming at a second location on the DNA template.

A second pyrimidine oligonucleotide used to prime the DNA polymerase I repair reaction was $C_3T_8$, which is present in two copies in the minus strand of S13 RF DNA. This experiment tested whether two different sites on the viral template could be primed simultaneously and investigated problems of determining two sequences in the same reaction mixtures.

The extension products of $C_3T_8$, separated by ionophoresis–homochromatography are shown in Figure 6. The schematic interpretation indicates that both sequences were extended and the sequence extension diverged after addition of one G residue onto each $C_3T_8$ primer. The $C_3T_7$ spot was due to degradation of the $C_3T_8$ primer by the 3′→5′ exonuclease activity of the DNA polymerase I. The spots marked a and b were examined by partial snake venom phosphodiesterase di-

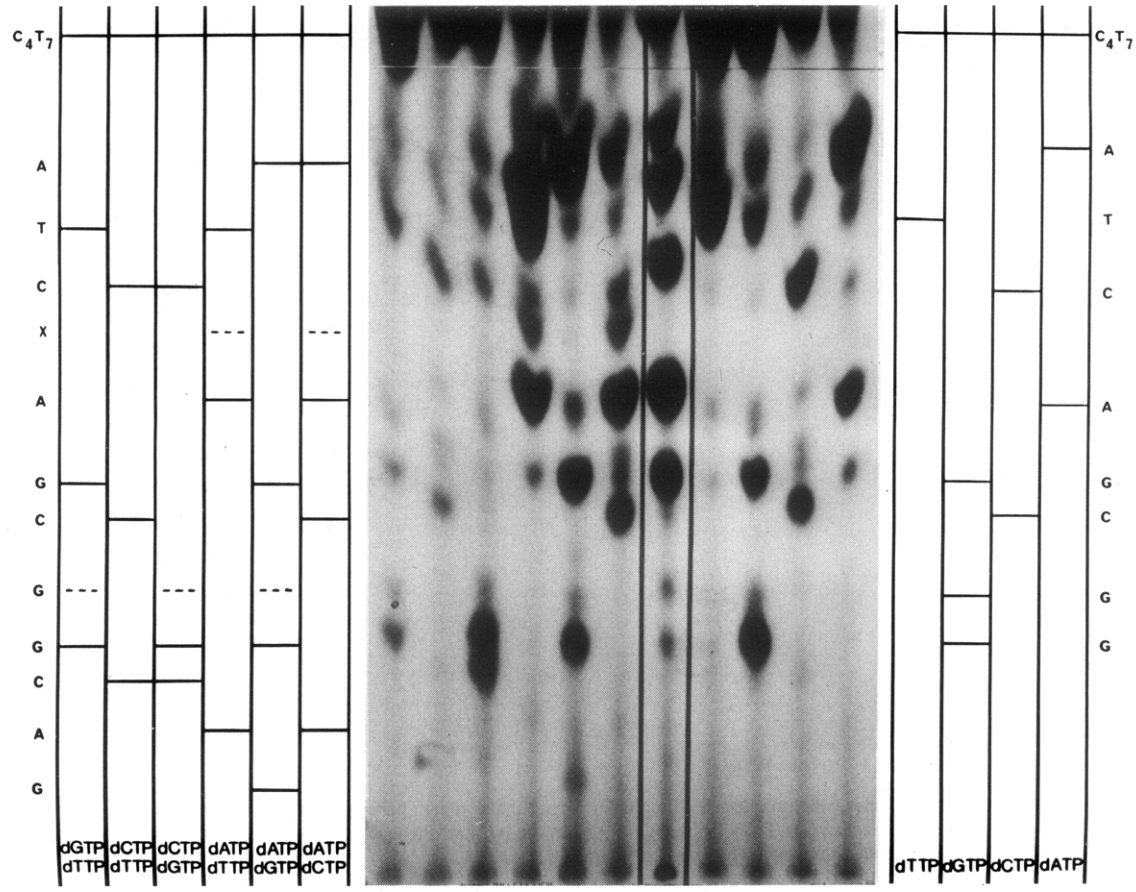


FIGURE 5: Unidimensional separation of the products resulting from the two stage extension of oligonucleotide primer $C_4T_7$. The products were separated by homochromatography on a 20 × 40 cm DEAE cellulose thin-layer plate at 60 °C, eluent 2% partially hydrolyzed yeast RNA containing 7 M urea. The separated oligonucleotide extension products were visualized by autoradiography. The separation in the center channel, bordered by lines, is of a primer extension reaction similar to that in Figure 3. The four channels to the right of center are the second stage extension in the presence of one dNTP only (see schematic for interpretation). The six channels to the left of center are the second stage extension in the presence of pairs of dNTPs (see schematic for interpretation).

TABLE IV: Sequence of Some Extension Products of $C_3T_8$.

| Oligonucleotide (from Figures 7 and 8) | Sequence | |
|---|---|---|
| | 5′ | 3′ |
| i | . . . . . T-T-T-G-C-G | |
| h | . . . . . T-T-T-T-T-G-C | |
| g | T-C-T-C-T-C-T-T-T-T-T-G | |
| $C_3T_8$ | T-C-T-C-T-C-T-T-T-T-T | |
| | C-C-T-T-T-C-T-T-T-T-T | |
| a | C-C-T-T-T-C-T-T-T-T-T-G-G-G-G | |
| b | . . . . . G-G-T-A | |
| c | . . . . . T-A-A | |
| d | . . . . . A-T-G-A-T | |
| e | . . . . . T-G-A-T-A | |
| f | . . . . . A-C-T-C-A-T | |

gestion and shown not to be derivatives of $C_3T_8$ primer extension or $C_3T_8$ 3′→5′ degradation. Their origin is unknown. Determination of the total sequence was by redigestion of selected extension products. The autoradiographs in Figure 7 are of partial snake venom phosphodiesterase digests of extensions a to f and showed that the $C_3T_8$ primer with the sequence C-C-T-T-T-C-T-T-T-T-T was extended by G-G-G-G-T-A-A-T-G-A-T-A-C-T-C-A-T-C. Digest d served to clarify some uncertainty in the assignments made of the three close spots, sequence –A-T-G– in Figure 6. Digest f was particularly important in deciphering the positions which are drawn dotted in Figure 6 and could not be read by mobility shifts alone from the original extension. This underlines the importance of the back digests for confirming sequences particularly where intermediate extension products may be absent or difficult to observe. The sequences from extensions a to f are listed in Table IV and each sequence overlapped the previous one. Autoradiographs of partial snake venom digests of extensions g, h, and i are shown in Figure 8. The $C_3T_8$ primer with sequence T-C-T-C-T-C-T-T-T-T-T was extended by –G-C-G–. The back digest of g confirms that both $C_3T_8$ primers are extended by one G residue. Since back digest a located one extension en-

TABLE V: Sequences of Extensions of Pyrimidine Oligonucleotide Primers in the DNA Polymerase I Repair Reaction.[a]

| Primer | Sequences |
|---|---|
| $C_4T_7$ | $^{3'}$(T-C-A)G-A-C-G-G-C-G-A-C-T-A-*T-T-T-C-C-T-T-T-C-C-T*$^{5'}$<br>$^{5'}$(A-G-T)C-T-G-C-C-G-C-T-G-A-T-A-A-A-G-G-A-A-A-G-G-A$^{3'}$ |
| $C_3T_8$ | $^{3'}$(C)A-C-T-A-C-T-C-A-T-A-G-T-A-A-T-G-G-G-G-*T-T-T-T-T-C-T-T-T-C-C*$^{5'}$<br>$^{5'}$(G)T-G-A-T-G-A-G-T-A-T-C-A-T-T-A-C-C-C-C-C-A-A-A-A-A-G-A-A-A-G-G$^{3'}$ |
| $C_3T_8$ | $^{3'}$(T)G-C-G-*T-T-T-T-T-C-T-C-T-C-T*$^{5'}$<br>$^{5'}$(A)C-G-C-A-A-A-A-A-G-A-G-A-G-A$^{3'}$ |
| $C_3T_6$ | $^{3'}$(A-C-A-G-T-C-A-C-C-T-G-T-T)A-G-*T-C-T-T-T-C-T-C-T*$^{5'}$<br>$^{5'}$(T-G-T-C-A-G-T-G-G-A-C-A-A)T-C-A-G-A-A-A-G-A-G-A$^{3'}$ |
| $C_5T_4$ | $^{3'}$(T-G-T-C-T)C-A-A-C-C-G-T-A-C-*T-C-C-T-C-T-T-C*$^{5'}$<br>$^{5'}$(A-C-A-G-A)G-T-T-G-G-C-A-T-G-A-G-G-A-G-A-A-G$^{3'}$ |

[a] The primer sequences are italicized. Sequences in parentheses have not been confirmed by an independent method. Sequences underlined are terminator codons. The complementary (minus) strand is 3′ → 5′; the viral (plus) strand is 5′ → 3′.

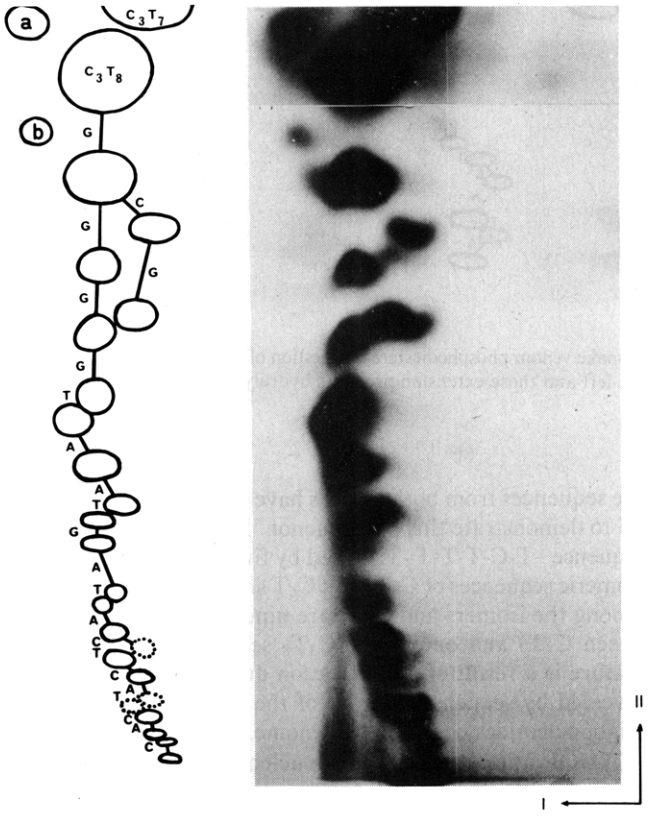


FIGURE 6: Two-dimensional fractionation of the 5′-$^{32}$P-labeled products resulting from extension of oligonucleotide isomeric primers $C_3T_8$ in the *E. coli* DNA polymerase I repair reaction, on an S13 viral DNA template. For details of the separation, see legend to Figure 1.

TABLE VI: Sequence Homologies among the Pyrimidine Oligonucleotides of S13 Minus Strand DNA and Viral DNA.[a]

| Oligonucleotide composition | Minus strand DNA sequences |
|---|---|
| $CT_6$ | T-T-T-T-T-T-C |
| $C_2T_7$ | T-T-T-T-C-T-T-C-T |
| $C_5T_4$ | C-T-T-C-T-T-C-T-C |
| $C_5T$ | C-C-C-C-T-C |
| $C_5T$ | T-C-C-C-C-C |
| $C_5T_3$ | T-C-C-C-C-T-T-C |
| $C_4T_7$ | T-C-C-T-T-T-C-T-T-T |
| $C_3T_8$ | C-C-T-T-T-C-T-T-T-T-T |
| $C_3T_8$ | T-C-T-C-T-C-T-T-T-T-T |
| $C_2T_7$ | T-C-T-C-T-T-T-T-T |
| $C_3T_6$ | T-C-T-C-T-T-T-C-T |
| $C_2T_6$ | C-T-C-T-T-T-T-T |
| | **Viral strand DNA sequences**[b] |
| $CT_6$ | C-T-T-T-T-T-T |
| $C_2T_8$ | C-T-T-T-T-T-T-C-T |
| $C_2T_6$ | C-T-T-T-T-C-T-T |
| $C_2T_6$ | T-T-T-T-C-T-T-C |
| $C_5T_6$ | C-T-T-C-C-T-C-T-C-T |
| $C_6T_4$ | C-T-C-C-T-C-T-C-C |
| $C_5T_4$ | C-T-T-C-C-T-C-C-T |
| $C_4T_4$ | C-T-T-C-C-T-T-C |
| $C_4T_5$ | C-T-C-T-T-T-C-T-C |
| $C_4T_5$ | T-C-T-T-T-C-T-C-C |
| $C_5T_3$ | C-C-T-T-T-C-C-C |
| $C_6T_3$ | T-C-C-T-T-C-C-C-C |
| $C_6T$ | C-T-C-C-C-C-C |
| $C_5T_2$ | C-T-C-T-C-C-C |

[a] Boxed areas indicate homologous sequences. [b] From Harbers et al. (1976).

tirely on one of the $C_3T_8$ primers (Figure 7) by inference the -G-C-G- must be on the remaining one. However, the possibility of misincorporation of residues on to the pyrimidine oligonucleotide which primed extensively cannot be excluded entirely at present. The two-step extension procedure was also used to confirm both sequences. The final confirmed sequences of $C_4T_7$ and $C_3T_8$ and sequence data using other pyrimidine oligonucleotide primers are listed in Table V.

## Discussion

The sequence analysis of the pyrimidine oligonucleotides from the minus strand completes the catalogue of long pyrimidine oligonucleotides present in S13 RF DNA once or twice. The abundance of true palindromes within the sequences, as shown in Table II, may prevent hairpin formation in the genome as discussed previously (Harbers et al., 1976) since they are inherently incapable of forming such hydrogen bonded structures. Another unusual feature of the sequences is the extensive internal sequence homology. This was observed also in the viral strand sequences (Harbers et al., 1976) and

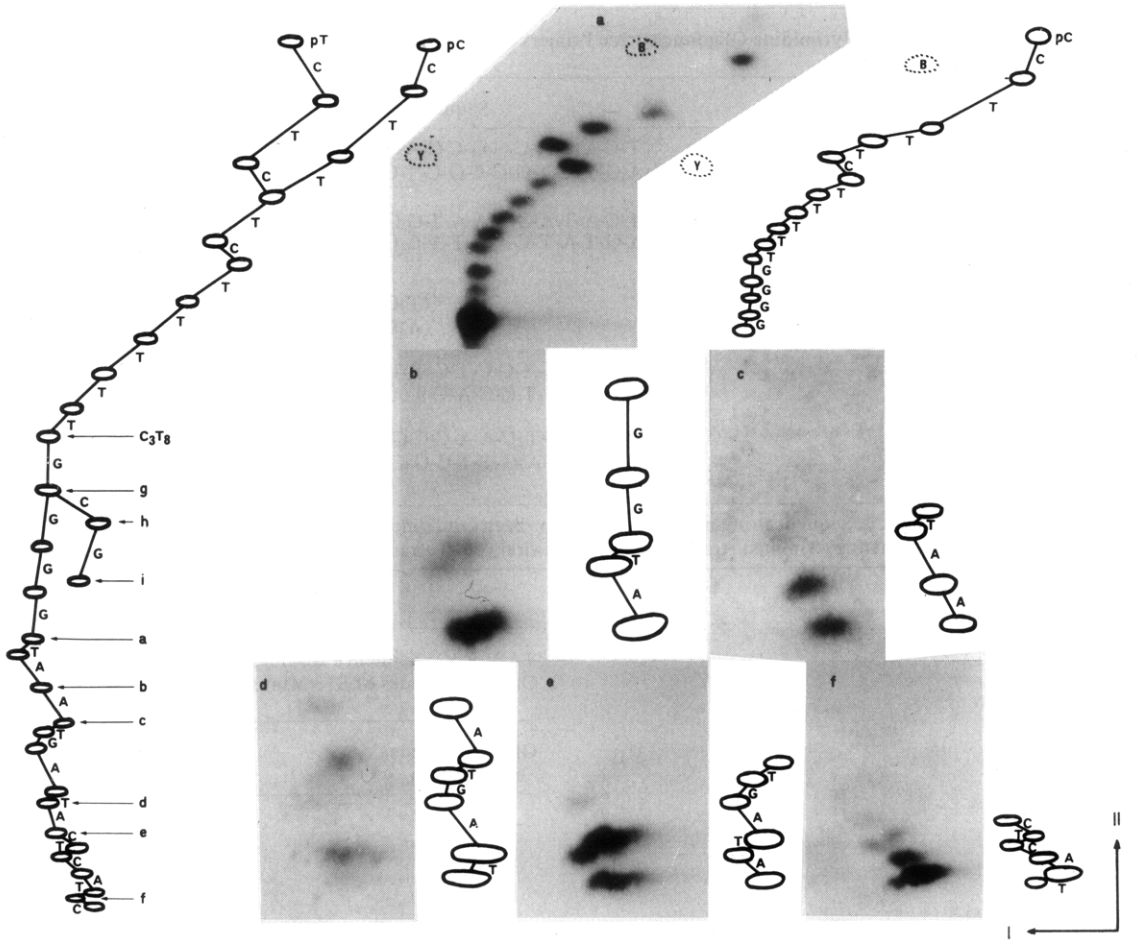


FIGURE 7: Two-dimensional fractionations of the products resulting from partial snake venom phosphodiesterase digestion of selected extension products of primer $C_3T_8$. The schematic of the extension of $C_3T_8$ from Figure 6 is on the left and those extension products hydrolyzed in this experiment are marked a to f inclusive. For details of the separation, see legend to Figure 1.

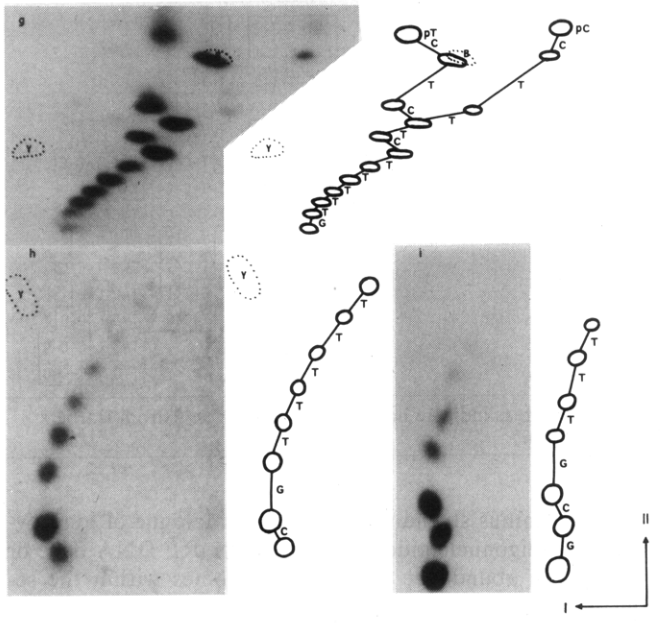


FIGURE 8: Two-dimensional fractionations of the products resulting from partial snake venom phosphodiesterase digestion of selected extension products of primer $C_3T_8$. The extension products hydrolyzed in this experiment are marked g, h, and i on the schematic on the left in Figure 7. For details of the separations, see legend to Figure 1.

the sequences from both strands have been arranged in Table VI to demonstrate this phenomenon. In the minus strand the sequence –T-C-T-T-T– is shared by five oligonucleotides. The isomeric sequences of $C_3T_8$ and $C_5T$ show extensive homology among the isomers and there are nine homologous bases between $C_4T_7$ and one of the $C_3T_8$ sequences. Whether this feature is a result of amplification during evolution may be revealed by sequence analysis of the regions in which these oligonucleotides occur in the genome.

The locations of those oligonucleotides, which have been used as primers in the DNA polymerase I repair reaction, on the *Hin*d and *Hae*III restriction fragment and genetic maps of S13 have been determined by pyrimidine cluster analyses of separated restriction enzyme fragments (Grosveld et al., 1976; Grosveld, 1977). The cumulative data are presented in Figure 9. $C_4T_7$ and $C_3T_6$ are in the gene H region. The $C_3T_8$ which extends to $C_3T_8G_4$ . . . is in the gene A region. The location was determined by the presence of $C_4$ in the corresponding restriction fragments (Grosveld, 1977). The other $C_3T_8$ isomer and $C_5T_4$ are in the same gene B, C region. A more exact location of these latter two primers will be forthcoming from experiments in progress based on the location of a single *Pst*I cleavage site in the gene B region (Goodchild and Spencer, 1978; Figure 9).

In Table V the viral strand sequences complementary to the primed sequences are presented since these are identical with the transcript sequences. The $C_4T_7$ primed sequence has one amino acid reading frame in which two sequential terminator

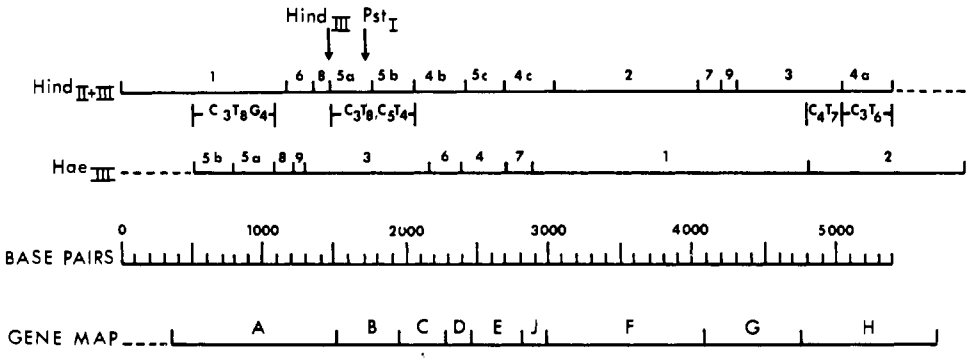


FIGURE 9: The position of the primers $C_4T_7$, $C_3T_8$, $C_3T_6$, and $C_5T_4$ on the genetic and physical map of bacteriophage S13. The physical map is a composite of data from Grosveld et al. (1976), Grosveld (1977), and Goodchild and Spencer (1978).

codons are present (see Table V). The location of $C_4T_7$ in gene H (Figure 9) indicates that this is not the reading frame used in vivo. The $C_3T_8$ primed sequence also has an amino acid reading frame in which two sequential terminator codons are present (Table V) and its location in gene A indicates it will not be a reading frame in vivo either. The imprecise location of $C_5T_4$ allows no speculation concerning the terminator codon nor the initiator codon superimposed by one frame, present in the primed sequence.

The present study shows clearly that pyrimidine oligonucleotides can be used as specific primers for the DNA polymerase I repair reaction. The experiment with the isomeric $C_3T_8$ oligonucleotide primers has demonstrated that oligonucleotides of chain length as short as 11 residues hybridize specifically to their complementary sequences and are extended by DNA polymerase I specifically and reproducibly. Prior determination of the sequence of isomeric primers allows separate identification of the different primed sequences and assignment to the correct primer. The experiment with the two $C_3T_8$ primers also revealed the superior priming of one isomeric oligonucleotide compared to the other. The reasons for this are not apparent but it may be an indication of secondary structure in specific regions of S13. Preliminary experiments in other systems have shown that pyrimidine oligonucleotides only 8 nucleotides long can be used as specific primers (Kaptein, 1977).

There are several advantages to the use of pyrimidine oligonucleotides as primers for the DNA polymerase I repair reaction. They are present in all DNAs studied, they can be isolated easily from any DNA, and they can be synthesized chemically. When restriction fragment maps of a DNA are available the oligonucleotides can be localized in the map by pyrimidine tract analysis (Grosveld et al., 1976; Grosveld, 1977). The small size of naturally occurring pyrimidine oligonucleotides eliminates the necessity of their removal from the primed sequence prior to the actual sequence analysis and chromatography of the primed product, and the immediately adjacent sequences are deduced also (cf. Sanger and Coulson, 1975).

This is the first report of oligonucleotides of chain length 9 to 11 residues, isolated from natural sources, being used as primers for sequence determination. Sanger et al. (1973, 1974) have used a synthetic octadeoxynucleotide A-C-C-A-T-C-C-A to prime a sequence in fd DNA by the ribosubstitution method. The primer contains purines which result in higher affinity for the template; nevertheless it primed at two sites on the template. Sanger and Coulson (1975) have used the purine rich synthetic decadeoxynucleotide A-G-A-A-A-T-A-A-A-A- as a primer in the plus-minus method and obtained extension of up to 40 nucleotides, on a $\phi$X174 DNA template.

The techniques developed for working with small pyrimidine oligonucleotides have several inherent advantages which can be applied to any small primers. The use of 5′-$^{32}$P-labeling eliminates the need for uniformly labeled material or [$\alpha$-$^{32}$P]dNTP substrates while maintaining high specific radioactivity in the final product. The procedure requires DNA polymerase I only, utilizing both its polymerase and 3′→5′ exonuclease activities. The Sanger and Coulson (1975) plus-minus procedure uses DNA polymerase I and T4 DNA polymerase. Based on the present study we have shown that T4 DNA polymerase can be replaced with DNA polymerase I in the plus-minus method (Darragh and Spencer, unpublished results). As noted above, with short primers there is no need to cleave the primer from the extended oligonucleotide prior to sequence analysis, a procedure that must be performed if restriction fragments are used as primers. This results in determination of the nucleotide sequence immediately adjacent to the primer. The basic approach is similar to that of Sanger and Coulson's (1975) plus-minus method but differs in the use of pairs of dNTPs as a second part of the "minus" stage. This modification did have advantages in the present study for additional confirmation of sequences. Whether it will be of routine use cannot be predicted at this time since some problems with artifacts were encountered. Resolution of the primed sequences can be extended much further using polyacrylamide gels, and these experiments are currently in progress. This report is intended to show that short pyrimidine oligonucleotides can be used successfully as specific primers in the DNA polymerase I repair reaction to obtain sequence information.

## References

Brownlee, G. G., and Sanger, F. (1969), *Eur. J. Biochem. 11,* 395-399.

Burton, K. (1967), *Methods Enzymol. 12A,* 222-224.

Cerny, R., Cerna, E., and Spencer, J. H. (1969), *J. Mol. Biol. 46,* 145-156.

Delaney, A. D., and Spencer, J. H. (1976), *Biochim. Biophys. Acta 435,* 269-281.

Garen, A., and Levinthal, C. (1960), *Biochim. Biophys. Acta 38,* 470-483.

Goodchild, B., and Spencer, J. H. (1978), *Virology* (in press).

Goulian, M., Goulian, S. H., Codd, E. E., and Blumenfield, A. Z. (1973), *Biochemistry 12,* 2893-2901.

Grosveld, F. G. (1977), Ph.D. Thesis, McGill University.

Grosveld, F. G., Ojamaa, K. M., and Spencer, J. H. (1976), *Virology 71,* 312-324.

Harbers, B., Delaney, A. D., Harbers, K., and Spencer, J. H.

(1976), *Biochemistry 15,* 407-414.
Hewish, D. R., Smith, K. K., and Spencer, J. H. (1976), *Virology 74,* 363-376.
Jay, E., Bambara, R., Padmanabhan, R., and Wu, R. (1974), *Nucleic Acids Res. 1,* 331-353.
Jovin, T. M., Englund, P. T., and Bertsch, L. L. (1969), *J. Biol. Chem. 244,* 2996-3008.
Kamen, R. (1972), *Biochim. Biophys. Acta 262,* 88-100.
Kaptein, J. S. (1977), Ph.D. Thesis, McGill University.
Kaptein, J., and Spencer, J. H. (1976), *Proc. Can. Fed. Biol. Soc. 19,* 120.
Ling, V. (1972a), *Proc. Natl. Acad. Sci. U.S.A. 69,* 742-746.
Ling, V. (1972b), *J. Mol. Biol. 64,* 87-102.
Oertel, W., and Schaller, H. (1972), *FEBS Lett. 27,* 316-320.
Oertel, W., and Schaller, H. (1973), *Eur. J. Biochem. 35,* 106-113.
Petersen, G. B., and Reeves, J. M. (1969), *Biochim. Biophys. Acta 179,* 510-512.
Richardson, C. C., Schildkraut, C. L., Aposhian, H. V., and Kornberg, A. (1964), *J. Biol. Chem. 239,* 222-232.
Sanger, F., and Coulson, A. R. (1975), *J. Mol. Biol. 94,* 441-448.
Sanger, F., Donelson, J. E., Coulson, A. R., Kossel, H., and Fischer, D. (1973), *Proc. Natl. Acad. Sci. U.S.A. 70,* 1209-1213.
Sanger, F., Donelson, J. E., Coulson, A. R., Kossel, H., and Fischer, D. (1974), *J. Mol. Biol. 90,* 315-333.
Sanger, Fair, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, J. C., Hutchison, C. A., III, Slocombe, P. M., and Smith, M. (1977), *Nature (London) 264,* 687-695.
Schekman, R. W., Iwaya, M., Bromstrup, K., and Denhardt, D. T. (1971), *J. Mol. Biol. 57,* 177-199.
Shleser, R., Ishiwa, H., Mannes, B., and Tessman, E. S. (1968), *J. Mol. Biol. 34,* 121-129.
Spencer, J. H., and Boshkov, L. (1973), *Can. J. Biochem. 51,* 1206-1211.
Sulkowski, E., and Laskowski, M., Sr. (1971), *Biochim. Biophys. Acta 240,* 443-447.
Tate, W. P., and Petersen, G. B. (1974), *Virology 57,* 77-84.

# Multiple Thymine Dimer Excising Nuclease Activities in Extracts of Human KB Cells†

Kem H. Cook and Errol C. Friedberg*

ABSTRACT: Crude extracts of human KB cells grown in suspension culture contain enzyme activity that catalyzes the preferential excision of thymine-containing pyrimidine dimers from UV-irradiated *E. coli* DNA specifically incised adjacent to dimer sites. Fractionation of KB cell crude extracts reveals the presence of three such activities with distinct affinities for both DEAE-cellulose and phosphocellulose. One of the activities (activity B) is distinguished by its $s_{20,w}$ (2.6) and isoelectric point (9.0) from the other two (activities A and C) which have similar $s_{20,w}$'s (3.0-3.2) and isoelectric points (6.0). All three differ in their extent of stimulation by divalent cation and inhibition by NaCl or a sulfhydryl group inhibitor. These results indicate that multiple 5′ → 3′ dimer excision nuclease activities exist in human cells; however, there is as yet no direct evidence that these enzymes are functional in nucleotide excision repair in vivo.

The molecular mechanism of nucleotide excision repair of DNA, specifically the excision of thymine dimer-containing nucleotides from UV-irradiated DNA, has been extensively studied in prokaryote systems (for recent reviews, see Grossman, 1974; Friedberg, 1975; Grossman et al., 1975). In all cases reported, endonucleases catalyze hydrolysis of phosphodiester bonds 5′ with respect to dimers in DNA (Kushner et al., 1971; Minton et al., 1975; Braun et al., 1976). Subsequent exonucleolytically catalyzed dimer excision, therefore, requires degradation of DNA in the 5′ → 3′ direction. In studies with cell-free systems of *E. coli*, the 5′ → 3′ exonuclease activities associated with DNA polymerases I and III catalyze thymine dimer excision (Kelly et al., 1969; Friedberg & Lehman, 1974; Pawl et al., 1975; Livingston & Richardson, 1975). In addition, both exonuclease V (Tanaka & Sekiguchi, 1975) and exonuclease VII of *E. coli* (Chase & Richardson, 1974), neither of which are associated with DNA polymerases, promote dimer excision from specifically preincised DNA in vitro. Enzyme activities comparable to some of these have been reported in extracts of *M. luteus* (Kaplan et al., 1969, 1971; Hamilton et al., 1973), and, in phage T4 infected *E. coli*, two phage-coded dimer excising 5′ → 3′ exonucleases have been identified (Ohshima & Sekiguchi, 1972; Friedberg et al., 1974; Shimizu & Sekiguchi, 1976).

Studies on the enzymology of thymine dimer excision have also been carried out using mammalian cell-free systems (for recent review, see Friedberg et al., 1977). A number of investigators have reported endonuclease activities that discriminate between unirradiated and heavily UV-irradiated DNA (Bacchetti et al., 1972; Van Lancker & Tomura, 1974; Tomura & Van Lancker, 1975; Brent, 1975; Bacchetti & Benne, 1975; Duker & Teebor, 1975); to our knowledge, however, none of these enzymes have been shown to catalyze phosphodiester

† From the Laboratory of Experimental Oncology, Department of Pathology, Stanford University, Stanford, California 94305. *Received September 15, 1977.* Studies supported by Research Grants CA-12428 from the U.S. Public Health Service and NP-174 from the American Cancer Society, and by Contract No. EY-76-S-03-0326 with the U.S. Energy Research and Development Association. E.C.F. is the recipient of U.S. Public Health Service Research Career Development Award No. 5K04-CA 71005-03. K.H.C. was supported by Training Grant CA-09151 from the U.S. Public Health Service.

0006-2960/78/0417-0850$01.00/0 