Chemical Synthesis of a Primer and Its Use in the Sequence Analysis of the Lysozyme Gene of Bacteriophage T4*

(DNA sequencing/repair synthesis/hybridization)

R. PADMANABHAN†, ERNEST JAY, AND RAY WU

Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, New York 14850

Communicated by Leon A. Heppel, February 8, 1974

ABSTRACT We have developed a general approach for determining the nucleotide sequence of a gene, with the aid of a deoxyribonucleotide primer of defined sequence. The selection of the primer sequence was based on a short segment of mRNA sequence of T4 phage lysozyme. A tetradecadeoxynucleotide primer was chemically synthesized and its sequence verified by sequence analysis. This primer was found to bind to the single-stranded region of the exonuclease III-treated T4 DNA, and specific nucleotides were incorporated to its 3' end. The result indicated that this primer was bound to the expected location on the T4 DNA. Therefore, long sequences of the T4 lysozyme gene can now be determined from this specific starting point.

One of the new approaches proposed for the sequence determination of specific internal regions of DNA is the use of oligonucleotides of defined sequences as primers (1, 2) for the DNA polymerase I-catalyzed repair synthesis to provide labeled oligonucleotides which can be sequenced by established methods (3). The feasibility of this approach has been experimentally verified by using an octanucleotide, which is complementary to the left-hand cohesive end of 186 DNA, as primer for the repair synthesis. A sequence of eight nucleotides added to the C' end of the primer has been determined (4). Subsequently, several sequences from internal regions of DNA have been determined with the aid of primers (5–7). In this communication, this method is used to determine the sequence of a specific segment of T4 DNA coding for the lysozyme. The reason for choosing this system is that the complete amino-acid sequence of this enzyme is known (8); thus, DNA sequence information will be valuable for understanding the degeneracy in the genetic code. Moreover, Streisinger et al. (9) were able to derive an unique 17-nucleotide-long sequence of the mRNA coding for the T4 lysozyme by analysis of double-frameshift mutants and with the use of the genetic code (10, 11). We report here the chemical synthesis of a tetradecadeoxynucleotide, d(A-G-T-C-C-A-T-C-A-C-T-T-A-A), which corresponds to part of the mRNA sequence as shown in Fig. 1, except that dpT was substituted for puT. The exact sequence of this tetradecadeoxynucleotide was verified by direct sequence analysis and it was used as a specific primer in the sequence determination of the lysozyme gene beyond the 3' end of the primer.

The chemical syntheses of several shorter segments of oligodeoxynucleotides, which are complementary to part of the mRNA sequence shown in Fig. 1, have been reported (12, 13).

MATERIALS AND METHODS

T4 DNA. T4 DNA was prepared according to the previously published procedure (14, 15). For our purpose, T4 DNA obtained by phenol extraction of the T4 phage was found to be sufficiently pure for sequence analysis. Therefore, the sucrose density gradient purification step (15) was omitted.

The Enzymes. The spleen phosphodiesterase was a gift of Dr. G. Bernardi. Escherichia coli exonuclease III and DNA polymerase I were purified according to Jovin et al. (16). Deoxynucleotidyl terminal transferase was a gift of Drs. Roychoudhury and Kössel.

Methods. Chemical synthesis of the tetradecadeoxynucleotide (14-mer) was according to the general methods developed by Khorana et al. (17–19). A summary of conditions for the synthesis of the tetradecanucleotide and the yield of the oligonucleotides is given in Table 1.

Exonuclease III Treatment of T4 DNA. The procedure is as described previously (7), except that the extent of digestion was 26%.

Hybridization of the Primer to DNA. An incubation mixture containing 0.4–0.6 pmols of exonuclease III-treated T4 DNA in potassium phosphate buffer (70 mM, pH 6.9) and 5 pmols of the primer was heated to 75° for 15 min and then cooled slowly to 45°. The reaction mixture was kept at 45° for 12 hr and then chilled to 5°. The solution was supplemented with 10 mM Mg++, 15 mM dithiothreitol, 60 mM NaCl, and 2–3 mM labeled deoxynucleoside triphosphates. The repair synthesis was started by the addition of 10 units of DNA polymerase I in a final volume of 0.44 ml. The unutilized primer and labeled deoxynucleoside triphosphates were separated from the DNA-oligonucleotide duplex by using a column (0.7 cm × 46 cm) of agarose (1.5 M, saturated with

Abbreviation: TEAB, triethylammonium bicarbonate.

* This is paper XV in a series on “Nucleotide Sequence Analysis of DNA.” Paper XIV is by R. Hamilton and R. Wu, J. Biol. Chem., in press.

† Present address: Institute for Molecular Virology, St. Louis University, School of Medicine, St. Louis, Mo. 63110.

Fig. 1. Sequence of a specific segment of mRNA coding for the T4 lysozyme. The mRNA sequence was derived from analysis of the amino acid sequences of wild-type and double-frameshift lysozyme mutants (9).
calf thymus DNA, 0.1 mg/ml, and washed free of the DNA). NH₄HCO₃ (0.05 M) containing 0.1 mM ethylenediaminetetraacetate (EDTA) was used as the elution buffer (1, 4). Fractions of 1.5 ml were collected and counted for ³²P by Cerenkov radiation in the liquid scintillation counter.

Dissociation of DNA-Oligonucleotide Duplex. After repeated evaporation to remove NH₄HCO₃, the DNA-oligonucleotide duplex was heated in 50 µl of water in a boiling water bath for 10 min. It was then fractionated as a line (1 cm) on one-dimensional homochromatography on DEAE-cellulose using partially hydrolyzed yeast RNA (Homo III). Alternatively, dissociation was achieved on a Sephadex G-100 column (0.5 cm × 63 cm) at 65° using 0.5 mM Tris-HCl buffer (pH 8.0 at 65°) as elution buffer.

3'-End-Group Analysis of Oligonucleotides. In the case when one-dimensional homochromatography was used for the dissociation of the DNA-oligonucleotide hybrid, the samples were scraped out of the DEAE-cellulose thin-layer plate using the Eppendorf plastic tips (W. Sarstedt, Inc.) with a cotton plug at the narrow end. A glass scraper (22) was attached at the wider end and the sample was collected in the tip by using a mild vacuum. Urea was removed from the sample by washing with approximately 2 ml each of 95% and 50% ethanol and then with water. The sample was eluted with 50-100 µl of 2 M triethylammonium bicarbonate (TEAB). After removal of the TEAB by evaporation, the dry sample was taken up in 20 µl of 0.05 M NH₄HCO₃ (pH 8.0), and the RNA carrier from the Homo-mix in the sample was digested with 4 µg of pancreatic RNase (Worthington) and 5 units of RNase T₁ for 8 hr at 37°. The sample was further digested for 1 hr at 37° with 0.04 µg of micrococcal nuclease in the presence of CaCl₂ (0.8 mM) and a mixture of four 3'-deoxyoligonucleotides and nucleosides (0.1 µmol each). Finally, spleen phosphodiesterase digestion was carried out in a volume of 0.05 ml which contained potassium phosphate buffer (25 mM, pH 6.0), ethylenediaminetetraacetate (1.4 mM), Tween 80 (0.04%), and the enzyme (0.25 µg). The incubation was carried out at 37° for 15 hr. The labeled mononucleotides and nucleosides were then separated and analyzed as described earlier (23).

RESULTS

It has been shown that the genes in T₄ phage are circularly permuted (24) and that the lysozyme gene is transcribed from the L-strand of T₄ DNA (25). This means that, statistically, the lysozyme gene is distributed along the entire L-strand length, in different DNA molecules. If 26% of the nucleotides are removed from the 3' end of the H-strand, it means that 26% of the total lysozyme gene has been rendered single-stranded, on a statistical basis. Thus, the maximum extent of hybridization of the synthetic tetradecanucleotide primer to the single-stranded template regions on the L-strand will be only 26%.

5'-End-Group Labeling of the Tetradecanucleotide for Its Sequence Analysis from Its 3' End and for Its Use as a Primer. The oligonucleotide was labeled at the 5' end with ³²P by polyuridylate kinase (26) in an incubation volume of 10 µl containing 200 pmoles of 14-mer, 66 mM Tris-HCl (pH 7.8), 6.6 mM MgCl₂, 15 mM dithiothreitol, 0.066 mM [γ-³²P]-ATP, and 5 units of the enzyme. The incubation was carried out for 2 hr at 37°. The labeled 14-mer was purified on a Sephadex G-50 column to remove the excess [³²P]ATP.

For sequence analysis, the labeled primer was digested partially with venom phosphodiesterase and fractionated by two-dimensional homochromatography (20-22). As shown in Fig. 2a, 14 spots were well resolved. From the characteristic mobility shifts of the oligonucleotides in our improved two-dimensional homochromatography system (7, 21, 32) and the knowledge that Spot 1 is [³²P]d[pa] by cochromatography and coelectrophoresis with dpA, the complete sequence of the tetradecamer was deduced to be 5'-³²P[d(pA-G-T-C-C-A-T-C-A-C-T-T-A-A)]-A.

For its use as a primer, the ³²P-labeled tetradecanucleotide was purified by one-dimensional chromatography on polyethyleneimine-cellulose thin-layer chromatography plate.

### Table 1. Summary of conditions and yields of condensation reactions

<table>
<thead>
<tr>
<th>3'-OH-containing component</th>
<th>(A) amount (µmoles)</th>
<th>5'-phosphate-containing component</th>
<th>(B) amount (µmoles)</th>
<th>TPS (µmoles)</th>
<th>Pyridine (ml)</th>
<th>Time (hr)</th>
<th>Yield (%)</th>
<th>Size of product^±</th>
</tr>
</thead>
<tbody>
<tr>
<td>d(mmt-bzA-A-OH)</td>
<td>2000</td>
<td>d(pmbG-OAc)</td>
<td>3140</td>
<td>6,000</td>
<td>25</td>
<td>5</td>
<td>68</td>
<td>dimer</td>
</tr>
<tr>
<td>d(mmt-bzA-mbG-OH)</td>
<td>1360</td>
<td>d(pT-OAc)</td>
<td>4000</td>
<td>10,000</td>
<td>20</td>
<td>5</td>
<td>67</td>
<td>trimer</td>
</tr>
<tr>
<td>d(mmt-bzA-mbG-T-OH)</td>
<td>930</td>
<td>d(panC-anC-OAc)</td>
<td>2400</td>
<td>5,400</td>
<td>15</td>
<td>7</td>
<td>43</td>
<td>pentamer</td>
</tr>
<tr>
<td>d(mmt-bzA-mbG-T-anC-anC-OH)</td>
<td>400</td>
<td>d(pbaA-T-OAc)</td>
<td>1500</td>
<td>4,500</td>
<td>10</td>
<td>7</td>
<td>10</td>
<td>heptamer</td>
</tr>
<tr>
<td>d(mmt-bzA-mbG-T-anC-bzA-T-OH)</td>
<td>41</td>
<td>d(pC-bzA-OAc)</td>
<td>600</td>
<td>1,440</td>
<td>5</td>
<td>6.5</td>
<td>32</td>
<td>nonamer</td>
</tr>
<tr>
<td>d(mmt-bzA-mbG-T-anC-anC-bzA-T-anC-bzA-OH)</td>
<td>8</td>
<td>d(pT-T-bza-bza-OAc)</td>
<td>100</td>
<td>600</td>
<td>2</td>
<td>6</td>
<td>31</td>
<td>14-mer</td>
</tr>
</tbody>
</table>

* Abbreviations and symbols used in this table: anC, N-anisoyldeoxycytidine; bzA, N-benzoyldeoxyadenosine; mbG, N-α-methylbutyryldeoxyguanosine; mmt, monomethoxytryptil; TPS, tri-isopropylbenzene sulfonyl chloride.

† The percent yield of all products is given as yield determined spectrophotometrically after complete purification by partition extraction or DEAE-cellulose column purification followed by precipitation.

‡ At the end of each step of the condensation reaction, a small amount of the protected compound was treated with NH₄OH followed by acetic acid to remove the protecting groups. The unprotected compounds were characterized after venom or spleen phosphodiesterase digestion and fractionation on paper chromatography (23). In every case, the expected ratios of each nucleotide and nucleoside were found.

§ The isolated yield was poor because a large amount of this protected heptamer was lost on the DEAE-cellulose column during the column purification step. This was due to the fact that the DEAE-cellulose used for this column had never been used before and probably a large amount of the heptamer was more or less irreversibly bound to the DEAE-cellulose. It was found that the loss of the precious long oligonucleotides can be minimized if protected mononucleotides or dinucleotides are first passed through the DEAE-cellulose column to saturate these irreversible binding sites. DEAE-cellulose which had been repeatedly used for the purification of protected oligonucleotides was found to give good recovery of protected long oligomers.
(cellulose impregnated with polyethyleneimine) using 1 M LiCl (27) and 7 M urea. This removed any nonphosphorylated starting material which would otherwise compete with the labeled 14-mer in the hybridization experiments with the T4 DNA template.

3'-End-Group Labeling of the Tetradecanucleotide for Its Sequence Analysis from Its 5' End and for Its Use as a Primer. The 14-mer was labeled at the 3' end with [32P]UTP and deoxynucleotidyl terminal transferase (28) in an incubation volume of 5 μl containing [32P]UTP (150 μM), Co++ ions (1.6 mM), dithiothreitol (16 mM), K-cacodylate (240 mM, pH 7.6), and 30 units of the enzyme. The incubation was

---

**FIG. 2.** (Legend appears at bottom of next page.)
Table 2. Binding of the synthetic 14-mer to exonuclease III-treated T4 DNA

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Components of binding reaction</th>
<th>Binding</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>T4 DNA + 14-mer</td>
<td>9</td>
</tr>
<tr>
<td>B</td>
<td>T4 DNA + 14-mer + polymerase</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>T4 DNA + 14-mer + polymerase +</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>dTTP</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>T4 DNA + 14-mer + polymerase +</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>dTTP + dGTP</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>T4 DNA + 14-mer + polymerase +</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>dTTP + dGTP + dGTP</td>
<td></td>
</tr>
</tbody>
</table>

The primer used was [5'-terminal-32P]dT(pA-G-T-C-C-A-T-C-A-C-T-T-A-A). Exonuclease III-treated T4 DNA (0.43 pmole) was hybridized with 5.2 pmole of [32P]14-mer as described under Materials and Methods. The dot over the labeled side of the labeled triphosphate denotes H-labeled compounds. The maximum possible extent of binding was 26% with this DNA.

The primer was used for a "monoaddition" product by treatment with alkali followed by bacterial alkaline phosphatase (28). The product was purified on a Sephadex G-50 column.

For sequence analysis by partial spleen phosphodiesterase digestion, the [32P]14-mer was prepared by oxidation of the 14-mer-[32P]dA by periodate followed by elimination (29). Fig. 25 shows the fractionation of a partial spleen phosphodiesterase digest of the [32P]14-mer by two-dimensional homochromatography. Spot 1 was shown to be [32P]dA and, from a series of successive additions of nucleotides causing characteristic shifts in mobility of spots 2 to 14, the complete sequence was deduced to be [3'-terminal-32P]dT(A-G-T-C-C-A-T-C-A-T-T-A-A).

The nucleotide next to the 14-mer at the 3' terminus was shown to be dP (Tables 2 and 3). Thus the transerase product of the 14-mer, 14-mer-[32P]pU (a 15-mer), was also used as a primer for hybridization to T4 DNA and repair synthesis. Since the conversion of the 14-mer to 14-mer-[32P]pU seemed to be almost quantitative, the purification step by polyethyleneimine-cellulose chromatography to remove the small amount of unutilized 14-mer was omitted.

### Table 3. Nearest-neighbor and 3' end group analyses of oligonucleotides after repair synthesis

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Primer used for repair synthesis</th>
<th>Products after enzyme digestion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[32P]14-mer + dTTP + Polymerase</td>
<td>dT, dA†, 14-mer-dT-OH</td>
</tr>
<tr>
<td>2</td>
<td>14-mer-[32P]pU-OH</td>
<td>dA†</td>
</tr>
</tbody>
</table>

Oligonucleotides after repair synthesis by DNA polymerase I were isolated by one-dimensional homochromatography and enzymatically digested as described under Materials and Methods. The nucleotides and nucleosides were separated and analyzed as described earlier (23).

The primer used was 14-mer-[3P]pU-OH (the 15-mer). Exonuclease III-treated T4 DNA (0.57 pmole) was hybridized with the primer (0.78 pmole) as described under Materials and Methods. The dot (·) over the nucleoside part of the labeled triphosphate denotes H-labeled compounds. The maximum possible binding was 26% with this DNA.

fact that the 15-mer (14-mer-[3P]pU) is also bound to T4 DNA and serves as a primer (see Table 4). Nearest neighbor analysis of the 14-mer-[3P]pU showed the transfer of 3P from [3P]pU to dpA, which was at the 3′ end of the 14-mer (Table 3, Experiment 2), as expected. In Experiment 3 of Table 3, the 14-mer-[3P]pU was extended to a 16-mer in the presence of [3H]dGTP and DNA polymerase I. The 16-mer showed a dG at the 3′ end and no dpG was found. Therefore, the sequence of nucleotides beyond the 3′ end of the chemically synthesized 14-mer is dp(T-G), which is in agreement with the RNA sequence as shown in Fig. 1.

DISCUSSION

The use of this general approach for DNA sequence determination is limited to the proper selection of the primer sequence which will hybridize to the single-stranded region of the template DNA of interest. Current methods in selecting the primer sequence include (a) the method based on knowledge of the RNA sequence, such as tRNA (8) or mRNA (this communication) and (b) the method based on knowledge of the protein sequence coded by the genome of interest, such as the endolysin gene from λ phage, and the correct choice of the triplet in cases where the amino-acid code is degenerate (1, 7).

The method used in this communication for choosing the primer sequence has the potential advantage that the DNA sequence information obtained beyond the primer would reveal the frequency of the actual use of the triplets among the degenerate codewords. This is important in understanding the degeneracy in the genetic code. The knowledge of the exact DNA sequence which codes for a protein is also helpful in the elucidation of the tertiary structure of its mRNA transcript. Such valuable information cannot be obtained in cases where the primer is bound to the DNA at an unknown location (5).

Once the primer is synthesized, a specific ribonucleotide can be added to its 3′ end by the deoxyribonucleotidyl terminal transferase for use as a more versatile primer. Such a primer, the 14-mer-[3P]pU for example, has an added advantage in the determination of longer sequences beyond its 3′ end because the oligonucleotides incorporated beyond the 3′ end of pU (in polymerase I-catalyzed repair synthesis) can be cleaved off from the primer with pancreatic ribonuclease, thus making sequence analysis of the newly synthesized segment easier.

The chemical synthesis of the tetradecanucleotide was carried out in the laboratory of Dr. H. G. Khorana between March and July of 1972. We are indebted to Dr. H. G. Khorana for his helpful advice and generous support; we also thank Dr. K. Agrawal for expert guidance. This work has been supported by grants GM-18887 from the National Institutes of Health and GB-40036X from the National Science Foundation.