Sequencing of large double-stranded DNA using the dideoxy sequencing technique

G. F. HONG

Medical Research Council Laboratory of Molecular Biology,
Hills Road, Cambridge CB2 2QH, U.K.

(Received 23 September 1982)

The dideoxy sequencing technique has been applied to the direct sequencing of large double-stranded DNA molecules with a small single-stranded primer. For instance, the method was applied to the lambda genome, which contains 48,502 base-pairs (Sanger F, Coulson AR, Hong GF, Hill D & Petersen GB, 1982, J. Mol. Biol., in press), and the coding region for gene W identified. The procedure proves useful in the sequence analysis of a large number of different mutations in a particular region and in the analysis of eukaryotic DNA cloned in plasmids, phages, and cosmids.

Normally the dideoxy sequencing technique uses a single-stranded DNA as the template (1). The use of single-stranded phage M13 vectors with this system has made it possible to determine the sequence of large molecules quickly (2,3).

Recently the M13 reverse-sequencing method (4) and the sequencing of pBR322 DNA (5-7) have shown that it is possible to sequence small double-stranded DNA molecules. However, such procedures have not been successful with large double-stranded DNA molecules. This paper demonstrates that it is possible to sequence directly large double-stranded DNA molecules using the dideoxy sequencing technique with a small single-stranded primer. This method was, for example, applied to the sequencing of the bacteriophage-lambda genome, which contains 48,502 base-pairs (8).

Recent progress in the field of solid-phase oligonucleotide synthesis means that primer synthesis is no longer a practical limitation. Starting with blocked monomers, a 20-base-long oligonucleotide can be synthesized and purified in a week. The method described here should therefore be useful for sequencing particular regions of the genome by using synthetic primers complementary to sites near regions of interest.

The procedure was used to identify the coding region for gene W in lambda by comparing the sequence obtained from a strain containing an amber mutant in gene W with the wild-type sequence. This procedure makes it possible to obtain sequence information directly from the genomic library DNA without sub-cloning.

* Permanent address: Shanghai Institute of Biochemistry, Chinese Academy of Sciences, Shanghai, China.
Materials and Methods

Materials

DNA polymerase I (large fragments) was obtained from Boehringer, Mannheim; dNTPs, ddNTPs, and T4 polynucleotide kinase from P.L. Biochemicals Ltd.; and $[^{32}P]dATP$ (sp. act. > 400 Ci/mmol) and $[^{32}P]ATP$ (sp. act. 3000 Ci/mmol) from Amersham International Ltd. The 20-nucleotide primer was prepared in this laboratory (9). Lambda bacteriophage c1857Ssm7Wam403 was a gift from Dr. R. Hendrix and the DNA prepared from this phage from Dr. D. Hill.

Preparation of 5'-end-labelled primer

100 µl of $[^{32}P]ATP$ (0.1 mCi) was dried in a siliconized tube and the label dissolved in a mixture of 14 µl of 20-nucleotide primer (1.6 pmol/µl), 2 µl of 0.5 M Tris/HCl, pH 7.5, 0.1 M MgCl$_2$, and 2 µl of 0.1 M DTT. The mixture was drawn into a capillary and 2 µl of T$_4$ polynucleotide kinase (approx. 10 units/µl) added. The capillary was sealed and incubated at 37°C for 1 h. The reaction mixture was diluted with 80 µl of 10 mM Tris/HCl, pH 7.4, 0.1 mM EDTA (buffer A) and extracted with 50 µl of phenol equilibrated with buffer A. The 5' $^{32}P$-labelled primer was precipitated with 3 vol. of 95% ethanol, 0.5 M NaCl, and the pellet was collected by centrifugation for 10 min at 12 000 g; it was then dissolved in 42 µl of water.

Sequencing procedure

The sequencing reaction was carried out in 5% glycerol and a high-pH/low-salt buffer (4 mM Tris/HCl, pH 8.3, 4 mM MgCl$_2$) as follows: 1 µl of lambda phage DNA (1 µg/1 µl) was mixed with 1 µl of labelled primer and 7 µl of water. The solution was sealed in a capillary, incubated in a boiling water bath for 3 min, and immediately transferred to a dry-ice/ethanol bath as described in ref. 4. 1 µl of 100 mM Tris/HCl, pH 8.3, and 100 mM MgCl$_2$ was added. The capillary was re-sealed and incubated at 50°C for 30 min. 2-µl samples of primer-template mixture were transferred into Eppendorf tubes and then combined with 1 µl of one of the following terminator mixtures (values shown are µl):

<table>
<thead>
<tr>
<th>Terminator Mixture</th>
<th>dC</th>
<th>dG</th>
<th>dA</th>
<th>dT</th>
</tr>
</thead>
<tbody>
<tr>
<td>dCTP (0.5 mM)</td>
<td>0.7</td>
<td>7.0</td>
<td>7.0</td>
<td>7.0</td>
</tr>
<tr>
<td>dGTP (0.5 mM)</td>
<td>7.0</td>
<td>7.0</td>
<td>7.5</td>
<td>7.5</td>
</tr>
<tr>
<td>ddC</td>
<td>0.7</td>
<td>7.0</td>
<td>7.0</td>
<td>7.0</td>
</tr>
<tr>
<td>ddG</td>
<td>7.0</td>
<td>7.0</td>
<td>7.5</td>
<td>7.5</td>
</tr>
<tr>
<td>ddA</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td>ddT</td>
<td>7.0</td>
<td>7.0</td>
<td>7.5</td>
<td>7.5</td>
</tr>
</tbody>
</table>

DNA polymerase I (large fragment, 5 units/µl) was diluted 40 times with 12.5% glycerol. 2 µl of the diluted enzyme was added to each reaction mixture. The reaction was carried out at room temperature for 15 min, followed by incubation at 50°C for a further 15 min.
Each reaction was then combined with 2 μl of 96% formamide, 10 mM EDTA, pH 7.0, 0.3% xylene cyanol FF, 0.3% bromophenol blue, and heated to 100°C for 3 min. 2.5-μl samples of each reaction were loaded onto a 6% polyacrylamide gel and electrophoresed. The gel was fixed and dried in a dual-temperature slab-gel dryer. Exposure was usually for 2 days at room temperature.

Results and Discussion

DNA sequence

The sequence obtained by this method can be read as far as the sequences obtained by the single-stranded M13 sequencing system, with low background and no ambiguities. Fig. 1 shows an autoradiograph of a 6% polyacrylamide/7 M urea sequencing gel giving the sequence extending from the 20-nucleotide primer into the W gene on the lambda genome.

![ Autoradiograph of 6% polyacrylamide/7 M urea sequencing gel, obtained by direct sequencing of lambda genome, giving the sequence of gene W with the amber mutation. W_i and W_t designate the initiation and termination codons respectively of the W gene. Thymine at position 26429 as indicated by the arrow, has turned the glutamine codon CAG into the termination codon TAG (see Fig. 2) by the amber mutation. Numbers indicate the nucleotide order on the lambda genome (8). ](image)
Fig. 2. Part of the lambda-phae cI857Sam7 sequence (8) covering the two possible reading frames (W1 and W2) for the W gene and part of the reading frame for the B gene. Cytosine at position 2642 has been changed into thymine (CAG → TAG) by the amber mutation (see Fig. 1), indicating that W1 is the correct reading frame for the W gene.
lambda genome cI857Sam7Wam403. The primer was designed to be positioned before the two possible reading frames for the \( \bar{w} \) gene. The sequence obtained from this lambda genome showed that the amber mutation has occurred in the first reading frame (Figs. 1 and 2). Therefore the correct coding region for gene \( \bar{w} \) has been determined.

**Primer**

The primer should be located about 15 nucleotides away from the start of the region of interest because this part cannot be read on the sequencing gel (data not shown). Shorter primers (13-mer and 14-mer) used in preliminary experiments produced ambiguous sequencing ladders. The SEQFIT computer programme (10) showed that in the case of the lambda genome longer primers were required to find a specific priming site within a limited space before the region of interest. 20-nucleotide primers were found to meet this requirement easily.

**Primer annealing and sequencing reaction**

Initial experiments using methods previously described gave very poor results. Rather specific conditions are required to achieve satisfactory results. Presumably it is necessary for two lambda strands to remain denatured while the primer anneals to the template. Annealing at room temperature (4,6,7) produces very weak sequencing patterns, which are masked by the presence of artifact bands. However, raising the annealing temperature to 60\(^\circ\)C for the primer-template combination (5) gives no sequencing patterns at all. It was found that annealing at 50\(^\circ\)C improved the sequencing patterns on the long running gel, but about the first 100 nucleotides from the priming site were still unreadable. It was assumed that complicated partial secondary structures may exist in the denatured double-stranded genome and prevent the polymerase from moving farther. Additional incubations of the reaction mixture at 50\(^\circ\)C, as described above, solved this problem. Addition of glycerol aids maintenance of enzyme activity at the higher temperature.

The concentration of the lambda DNA in the reaction mixture is important. It must not exceed 0.003 pmol/\( \mu \)l or the sequencing reaction will be inhibited, probably due to the re-annealing of the strands. The concentration of the template DNA used here was approximately 0.001 pmol/\( \mu \)l, which was one-tenth of that usually used in the method (6,7).

When the label was incorporated by the normal procedure as [\( \alpha ^{32}P \)] triphosphate during the sequencing reaction, the gels gave very high background such that many bands could barely be recognized; this was probably due to 'filling in' by the enzyme at damaged sites on the lambda genome. Therefore it is necessary to use the end-labelled primer.

In summary, it has been demonstrated in this work that the DNA sequence can be quickly and easily determined by the dideoxy method on the double-stranded large DNA molecules using synthetic primers. This should prove useful in the sequence analysis of a large number of different mutations in a particular region (e.g. the Y region
of phage lambda [11]) and in the analysis of eukaryotic DNA cloned in plasmids, phages, and cosmids.

Acknowledgements

The author would like to thank Dr. F. Sanger for his encouragement and advice during the course of this work and for a critical reading of the manuscript. The gift of lambda c1857Sam7Wam403 DNA from Dr. D. Hill is also appreciated. The author has been given leave of absence by the Chinese Academy of Sciences and is a temporary member of the Medical Research Council's scientific staff.

References