DNA Sequencing by Indexer Walking

KATARZYNA GROMEK and TADEUSZ KACZOROWSKI*

Background: There is a need for DNA sequencing methods that are faster, more accurate, and less expensive than existing techniques. Here we present a new method for DNA analysis by means of indexer walking. Methods: For DNA sequencing by indexer walking, we ligated double-stranded synthetic oligonucleotides (indexers) to DNA fragments that were produced by type IIS restriction endonucleases, which generate nonidentical 4-nucleotide 5' overhangs. The subsequent amplification (30 thermal cycles) of indexed DNA provided a template for automated DNA sequencing with fluorescent dideoxy terminators. The data gathered in the first sequencing reaction permitted further movement into the unknown nucleotide sequence by digestion of analyzed DNA with selected type IIS restriction endonuclease followed by ligation of the next indexer. A library of presynthesized indexers consisting of 256 oligonucleotides was used for bidirectional analysis of DNA molecules and provided universal primers for sequencing.

Results: The proposed protocol was successfully applied to sequencing of cryptic plasmids isolated from pathogenic strains of Escherichia coli. The overall error rate for base-calling was 0.5%, with a mean read length of 550 nucleotides. Approximately 1000 nucleotides of high-quality sequence could be obtained per day from a single clone.

Conclusions: Indexer walking can be used as a low-cost procedure for nucleotide sequence determination of DNA molecules, such as natural plasmids, cDNA clones, and longer DNA fragments. It can also serve as an alternative method for gap filling at the final stage of genome sequencing projects.

© 2005 American Association for Clinical Chemistry

In the genomic era, DNA sequencing has emerged as a leading diagnostic method in clinical laboratories (1, 2).

The original procedure based on the use of dideoxy terminators (3) has undergone many improvements, including the development of alternative approaches (e.g., pyrosequencing and sequencing by hybridization), making DNA sequencing feasible and indispensable in molecular diagnosis of hereditary disorders, detection of somatic mutations, and tracing of single-nucleotide polymorphisms. One widely used strategy for DNA sequencing is based on analysis of large DNA molecules by primer walking (4). This strategy makes use of individual primers specifically designed for each sequencing step and enables direct and systematic analysis of DNA fragments on both strands with low redundancy and without subcloning. The primer-walking strategy is attractive for large-scale projects, but primers are expensive and delays are associated with their design and synthesis.

Recent progress in sequencing of many genomes was achieved with the random shotgun approach (5). However, the aforementioned drawbacks of primer walking could be bypassed by the use of (a) presynthesized primers [e.g., 8mers (6), 9mers (7), or 10mers (8)], (b) methods for primer assembly from 5- or 6mer libraries (9–14), or (c) primers produced by high-throughput multichannel oligosynthesizers (15, 16). Unfortunately, the use of these methods is limited because of difficulties connected with processing and handling large libraries of oligonucleotides or the sophisticated equipment needed for their synthesis.

Our approach to DNA sequencing by primer walking bypasses the requirement for custom primer synthesis and relies on a universal set of 256 presynthesized oligonucleotide adaptors, referred to as indexers (17), and type IIS restriction endonuclease that cleave DNA at defined distances from their recognition sites, generating ambiguous 4-nucleotide (nt) 5' overhangs (18).

Materials and Methods

BACTERIAL STRAINS
We used Escherichia coli E1585-68 (19), E. coli 904, and E. coli 278. Strains 904 and 278 were isolated from patients admitted to Nicolaus Copernicus Hospital (Gdańsk, Poland). Bacteria were cultivated at 37 °C in Luria-Bertani medium (20).
Fig. 1. Schematic of DNA sequencing by indexer walking.
For details, see the Materials and Methods section.
DNA sequencing by indexer walking

Method principle. A schematic representation of DNA sequencing by indexer walking is shown in Fig. 1. Before sequencing, DNA is cloned into vector to obtain a genomic library with inserts of 4- to 5-kb length. The protocol for DNA sequencing by indexer walking consists of 3 consecutive steps: (a) DNA is amplified by PCR, followed by DNA sequencing with the universal primer (M13/pUC forward or reverse primer, or the common indexing primer; Table 1); (b) the obtained nucleotide sequence is then analyzed with a computer program to find specific sites recognized by type IIS restriction endonucleases located close to its 3’ end and to determine the composition of the 4-nt protruding ends produced by each enzyme; and (c) analyzed DNA is partially digested with a selected type IIS restriction endonuclease. This step is followed by ligation of a specific indexer, complementary to the cut site determined in a previous step, to one of the products of partial digestion. In each round of indexing, DNA processed in the previous cycle is used. Finally, the common primer provided by the indexer is used in a DNA-sequencing step in which reaction products are analyzed by capillary electrophoresis with an automated sequencer. Subsequent repetition of these steps permits unidirectional movement into the unknown sequence because this method is based on the data received in each round of DNA indexing.

Indexer assembly. Indexers used in this study were double-stranded adaptors composed of a 20-nt common indexer primer annealed to 1 of the 256 individually synthesized 24-nt indexer strands with 4-nt 5’-protruding ends complementary to the 4-nt nonidentical 5’ overhangs produced by selected type IIS restriction endonucleases. The total possible number of such ends is 256 (4^4). The library of indexers therefore consists of 256 individual adaptors, which differ from each other by 4 nucleotides at the 5’-protruding end of the indexing strand. The oligonucleotides used in the experimental procedures (Table 1) were supplied by the Institute of Biochemistry and Biophysics PAN (Poland) or Interactiva (Germany) in a desalted form and were used without additional purification. Indexing strands (24mers; Table 1) were phosphorylated at the 5’ terminus to make them suitable for ligation to DNA fragments produced by type IIS restriction endonucleases.

Each indexing strand (1 nmol) was phosphorylated for 30 min at 37 °C in a 10-μL reaction mixture containing polynucleotide kinase buffer [70 mmol/L Tris-HCl (pH 7.6), 10 mmol/L MgCl₂, 5 mmol/L dithiothreitol], 1 mmol/L ATP, and 10 μL of T4 polynucleotide kinase (New England Biolabs). The reaction was stopped by heating for 2 min at 80 °C. The short oligonucleotide (20mer) that served as the indexer-specific primer for DNA amplification and sequencing (Table 1) was not phosphorylated at the 5’ terminus. For the annealing step, 500 pmol of appropriate 24mer and 20mer oligonucleotides were mixed in a 10-μL volume of 1× T4 DNA ligase buffer [33 mmol/L Tris-acetate (pH 7.8), 66 mmol/L potassium acetate, 10 mmol/L magnesium acetate, 0.5 mmol/L dithiothreitol], heated at 65 °C for 5 min, and then allowed to cool slowly to room temperature. The concentration of each indexer stock solution was 10 μmol/L.

Template preparation. We isolated plasmid pEC157 from E. coli E1585-68 by the alkaline/sodium dodecyl sulfate extraction procedure (20). Plasmid pHYB157 was constructed by cloning plasmid pEC157 linearized with PvuII enzyme into vector pGEM3Zf(+) (Promega) digested with Hincll. Plasmid pHYB157 was used as a template in DNA sequencing. In the first step, M13/pUC universal

<table>
<thead>
<tr>
<th>Table 1. Oligonucleotides used in this study.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oligonucleotide</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Ind1</td>
</tr>
<tr>
<td>Ind2</td>
</tr>
<tr>
<td>Ind3</td>
</tr>
<tr>
<td>Ind4</td>
</tr>
<tr>
<td>Ind5</td>
</tr>
<tr>
<td>Ind6</td>
</tr>
<tr>
<td>Ind7</td>
</tr>
<tr>
<td>Ind8</td>
</tr>
<tr>
<td>Ind9</td>
</tr>
<tr>
<td>Ind10</td>
</tr>
<tr>
<td>Common indexer primer (20mer)</td>
</tr>
<tr>
<td>M13/pUC universal primers (20mers)</td>
</tr>
<tr>
<td>Forward (−23)</td>
</tr>
<tr>
<td>Reverse (−28)</td>
</tr>
</tbody>
</table>

^a Underlined bases in the indexing strands indicate the 4-nt 5’ overhangs complementary to the cut site produced by a type IIS restriction endonuclease and responsible for indexer specificity.

^b pEC157 DNA sequence has been filed with GenBank as accession no. AF432497.
primers (–23 forward and –28 reverse; Table 1) were used in the PCR. The product of DNA amplification was isolated from agarose after gel electrophoresis. Extracted DNA was sequenced by use of either M13/pUC –23 forward primer or M13/pUC –28 reverse primer. The obtained DNA sequence was analyzed to find restriction sites recognized by type IIS restriction endonucleases that generate ambiguous 4-nt overhangs. After that, the template was partially digested with the appropriate type IIS restriction enzyme at the 3' end of the sequenced region. The overall goal of this treatment was to make the template shorter after each step of automated DNA sequencing. Of 15 type IIS restriction endonucleases that are compatible with our method, 10 are commercially available (21). They recognize asymmetric sequences varying from 4 to 7 bp (4-bp cutter, SfiI; 5-bp cutters, Alw26I, BbvI, BsmFI, FokI, SfaNI, and Stsl; 6-bp cutters, AccIII, BbrI, BspMI, BtgZI, Eco31I, and Esp3I; 7-bp cutter, AarI). The DNA fragments produced by these enzymes typically are 128 to 8196 bp in length. Each partial digestion was performed in a 20-μL reaction mixture containing 300 ng of amplified DNA and 2 U of type IIS restriction endonuclease of choice (SfiNI, BbvI, or FokI; New England Biolabs). The reactions were carried out for 15 min at 37 °C in 1× reaction buffers provided by the manufacturer (New England Biolabs), followed by heat inactivation of the enzyme and phenol–chloroform extraction. After ethanol precipitation, the precipitate was collected by centrifugation in a microcentrifuge, dried briefly, and dissolved in 10 μL of 0.1× Tris-EDTA buffer [1 mmol/L Tris-HCl (pH 8.0), 0.1 mmol/L EDTA].

DNA indexing and amplification. Indexer ligation was performed in 10 μL of reaction mixture containing 1× T4 DNA ligase buffer supplemented with 1 mmol/L ATP, 60 ng of partially digested DNA, 10 U of highly concentrated T4 DNA ligase (Epicentre), and 100 fmol of the appropriate indexer (3-fold molar excess over 5' overhangs). Ligation was carried out for 1 h at room temperature, followed by heating at 65 °C for 10 min to inactivate the enzyme. When one of the M13/pUC primers was biotinylated, indexed DNA was recovered by mixing 5 μL of ligation mixture with 40 μL of 1× phosphate-buffered saline–washed streptavidin-coated beads (Streptavidin MagnaSphere Paramagnetic Particles; Promega). Shorter DNA molecules with the newly attached indexer were separated from the rest of the reaction mixture with a magnetic separation stand (Promega). After incubation at room temperature (5–10 min), the beads were washed twice with 2× binding and washing solution [10 mmol/L Tris-HCl (pH 7.5), 1 mmol/L EDTA, 2 mol/L NaCl] and then suspended in 20 μL of 10 mmol/L Tris-HCl (pH 8.0). Individual PCR reactions (50 μL total volume) contained 1× reaction buffer [10 mM Tris-HCl (pH 8.8 at 25 °C), 1.5 mM MgCl₂, 50 mM KCl, 1 mM/L Triton X-100], 0.8 mM deoxynucleoside triphosphates, 5 μL of bead suspension as template, 50 pmol of each primer [common indexer primer and M13/pUC forward (–23) or reverse (–28) primer; Table 1], and a mixture of 2 thermostable DNA polymerases: 4 U of DyNAzyme II DNA polymerase (Finzymes) and 0.2 U of Pfu polymerase (Fermentas). The reactions were run on an MJ Research PTC-100 thermocycler with the following cycling conditions: 95 °C for 2 min; 30 cycles of 94 °C for 40 s, 57 °C for 40 s, and 72 °C for 2 min; and a final step at 72 °C for 5 min. The reaction mixtures were then held at 4 °C until needed. In case of longer templates, the time needed for primer elongation in PCR reactions should be extended as recommended by others (22). After amplification, PCR products were separated by electrophoresis in 1% agarose gels, and fragments of the expected lengths were excised from the gel and purified by use of a DNA Purification Kit (A&A Biotechnology).

DNA sequencing. PCR-amplified DNA fragments were sequenced by the dideoxy termination method (3) with the ABI PRISM Big Dye Terminator Cycle Sequencing Ready Reaction Kit (Ver. 2.0) with AmpliTaq DNA Polymerase, FS (Perkin-Elmer, Applied Biosystems), and suitable primers (Table 1), as recommended by the manufacturer. DNA sequencing was performed in a reaction mixture containing 8 μL of Terminator Ready Reaction Mix (Perkin-Elmer, Applied Biosystems), 200 ng of the template, and 3.4 pmol of M13/pUC universal primer (forward or reverse) or common indexer primer in a total volume of 20 μL. Cycling was performed on an MJ Research PTC-100 thermocycler with the following thermal profile: 96 °C for 30 s, 50 °C for 15 s, and 60 °C for 4 min, repeated 25 times. The reaction products were precipitated with 80 μL of 750 mL/L isopropanol. After a 5-min incubation at room temperature, the DNA was recovered by centrifugation (25 min at 10 000g), and the supernatant was discarded. Precipitated DNA was washed with 200 μL of 750 mL/L isopropanol. The centrifugation step was repeated, and the sample was dried briefly at room temperature. The DNA pellet was then resuspended in 6 μL of deionized formamide and heated for 1 min at 94 °C. Just before the sample was loaded, 3 μL of DNA from the previous step was mixed with 20 μL of TSR reagent (Perkin-Elmer, Applied Biosystems) and denatured (94 °C for 3 min). Alternatively, unincorporated fluorescent dideoxy terminators were removed from the sequencing reaction by use of an Ex Terminator spin column according to the manufacturer's instructions (A&A Biotechnology). Sequencing products were separated on an ABI PRISM 310 Genetic Analyzer (Perkin-Elmer, Applied Biosystems) equipped with a 60-cm capillary filled with POP6 polymer, under standard running conditions (2.5 h). The obtained sequence data were analyzed by Vector NTI (Informax) and BLAST software (available through the National Center for Biotechnology Information at http://www.ncbi.nlm.nih.gov/BLAST/).
Results and Discussion

DNA sequencing by indexer walking relies on a set of 256 presynthesized adaptors that can be ligated to DNA fragments produced by type IIS restriction endonucleases that generate ambiguous 4-nucleotide 3’ overhangs. The method allows a particular fragment possessing a specific cohesive end to be isolated from a mixture of products obtained after partial digestion of target DNA, as depicted in Fig. 1.

Efficient and specific ligation of indexers is vitally important for the procedure. After trying T4 DNA ligases from different vendors, we decided to use highly concentrated enzyme (10 Weiss units/µL) provided by Epicentre. Several steps were taken to optimize the method. To suppress nonspecific ligation of indexers, a high ionic strength buffer was used. Others have shown that buffers with high ionic strength significantly improve specificity of ligation of double-stranded adaptors to DNA fragments digested by type IIS restriction endonucleases. Under these conditions, electrophoretic analysis of amplification products revealed that the nonspecific bands were rarely present (Fig. 2, A and B). The same effect can be accomplished by supplementing ligation mixtures with spermidine. We have also found that the most efficient ligation occurred at room temperature (22 °C).

Indexed DNA was amplified with M13/pUC forward or reverse primers and a common indexer primer. To facilitate the recovery of DNA templates, we modified the 5’ ends of primers (M13/pUC; Table 1) used in DNA amplification, thus enabling separation of shorter DNA molecules with the newly attached indexer, at the other end, from the rest of the reaction mixture (Fig. 1, step c). The only technical limitation that may impede the use of DNA sequencing by indexer walking is the length of the PCR-amplified DNA fragment. Our proce-

![Image](https://example.com/image1)

Fig. 2. Amplification of specific DNA fragments from plasmid pHYB157 by use of indexers.

This plasmid was constructed by cloning plasmid pEC157 linearized with PvuII enzyme into vector pGEM3Zf(+) digested with Hincll. Particular products were generated by use of an M13/pUC – 23 forward (A) or M13/pUC – 28 reverse primer (B) in conjunction with a specific indexer that was ligated to the cut site produced by the appropriate type IIS restriction endonuclease. The size of each fragment indicates the position of a type IIS restriction site with respect to the forward or reverse primer. Presented products were generated as mentioned above, with the following indexers: panel A, lane 2, ind1 (BbvI); lane 3, ind2 (FokI); lane 4, ind3 (BbvI); lane 5, ind4 (BbvI); lane 6, ind5 (SfaNI); panel B, lane 2, ind6 (SfaNI); lane 3, ind7 (BbvI); lane 4, ind8 (BbvI); lane 5, ind 9 (FokI); lane 6, ind10 (BbvI). In both panels, lane 1 is a product of DNA amplification performed with M13/pUC universal primers (– 23 forward and –28 reverse; lanes M1 and M2 contain molecular markers (λ/HindIII and pUC19/HinfI, respectively). Oligonucleotides used in the indexing procedure are listed in Table 1. (C), distribution of sites recognized by type IIS restriction enzymes producing ambiguous 4-nucleotide 3’ overhangs in the plasmid pEC157 linear form. The particular cut sites used for ligation of specific indexers are indicated, and their coordinates are given in parentheses. R, reverse; F, forward.
Determine works well with DNA fragments 4–5 kb in length; however, use of a mixture of 2 thermostable DNA polymerases, one of which has the 3’→5’ proofreading activity, permits amplification of longer fragments (up to 35 kb) more efficiently than use of only 1 polymerase without the 3’-exonuclease activity (22, 26, 27). In our experiments, we used a mixture of DyNAzyme II DNA polymerase from Thermus brockianus (Finzymes) and Pfu DNA polymerase from Pyrococcus furiosus (Fermentas), which possesses 3’→5’ proofreading activity. These enzymes were used at a ratio of 1:20. The amount of DNA received and its quality after PCR amplification (30 cycles) enabled automated DNA sequencing. The signal intensity was good (data not shown). Moreover, we have found that the PCR amplification step before DNA sequencing can be omitted by direct ligation of indexers to the products of partial digestion of the DNA of interest cloned in high-copy-number plasmids (data not shown). In our method, the overall base-calling error was only 0.5% with a mean read length of 550 nt. Movement in leaps of ~400–500 nucleotides is therefore feasible. Our procedure makes the analyzed molecule shorter after each round of indexing (Fig. 2, A and B), similar to the nested-deletion method (28); chances for amplification of secondary products are therefore decreased after each cycle. Moreover, DNA sequencing by indexer walking enables the analysis of both strands of the template; oversequencing can thus be kept low. We usually obtain ~1000 nt of high-quality sequence per day from a singular clone; therefore, sequencing of 4-kb DNA fragments in a highly systematic manner can take 4–5 days. Our method also reduces the costs associated with custom-primer synthesis. The estimated cost for an entire library of indexers is approximately US $4000.00 (the price for 256 individually synthesized oligonucleotides), and once purchased these can last for a long time because indexers are used at low concentrations. With the prices of synthetic oligonucleotides decreasing each year (29), the purchase of such a library is affordable for almost any laboratory dealing with DNA sequencing or molecular diagnostics.

DNA sequencing by indexer walking has been tested with plasmid pEC157 isolated from a pathogenic strain of E. coli (19). Plasmid pEC157 is a circular, extrachromosomal element carried by the pathogenic E. coli strain E1585-68, with 30 copies per bacterial cell. The complete nucleotide sequence of pEC157 has been determined on both strands (GenBank accession no. AF432497). We obtained the entire 2314-bp sequence (47.8% G+C) with the use of 5 indexers for 1 strand and 5 for the complementary strand (Table 1), with an overall redundancy of 2.9. The scheme for sequencing of pEC157 plasmid by indexer walking is depicted in Fig. 2C, which also shows the dense distribution of specific sequences recognized by type IIS restriction endonucleases producing 4-nt 5’ overhangs. With the same technology, we have also been able to sequence 2 other plasmids isolated from pathogenic strains of E. coli (GenBank accession nos. AY589570 and AY589571 for pEC904 and pEC278, respectively).

Other methods exploit features of DNA indexer technology for walking from known nucleotide sequences into unknown DNA regions. The movement is rendered possible by amplification of DNA fragments of interest with a specific primer complementary to the adjacent known sequence and with a second primer specific to the type II (30–33) or type IIS restriction endonuclease cleavage site (17, 22, 34–37). These methods are rarely applied for sequencing of large DNA fragments but are useful in reading only a few hundred nucleotides starting from the known sequence and moving in only one or in both directions (38). From this perspective, our indexer walking method seems to be the most advanced approach to indexer technology in DNA sequencing.

In conclusion, indexer walking is a low-cost method for nucleotide sequence determination of DNA molecules such as natural plasmids and cDNA clones and longer stretches of DNA, and is an alternative method for gap filling at the final stages of genome sequencing projects. Larger contigs can be assembled on the base of overlapping sequences; therefore, sequencing of a given DNA molecule can be carried out in a very systematic manner, without any gaps. Thus, the time needed to complete a sequencing project can be reduced substantially. Extensive DNA subcloning is avoided in indexer walking, making it simpler than other currently used sequencing strategies, such as whole-genome random shotgun sequencing. DNA sequencing by indexer walking may thus facilitate analysis of DNAs that exert potentially toxic effects or cannot be maintained in host cells (39–41).

We are grateful to Ania Kaczorowska for excellent help with the figures and editing the manuscript. This work was supported by University of Gdansk research grant BW/1170-5-0166-3. Sequence data from this report have been deposited with the GenBank Data Library under accession nos. AF432497, AY589570, and AY589571.

References


11. Azhikina TL, Veselevskaya SV, Myasnikov VA, Potapov VK, Ermoslayeva O, Sverdlov ED. Strings of contiguous modified pentamers with increased binding affinity can be used for DNA sequencing by primer walking. Proc Natl Acad Sci U S A 1993;90:11460–2.


