Genetic Applications of an Inverse Polymerase Chain Reaction

Howard Ochman, Anne S. Gerber and Daniel L. Hartl

Department of Genetics, Washington University School of Medicine, St. Louis, Missouri 63110

Manuscript received August 8, 1988
Accepted August 8, 1988

ABSTRACT

A method is presented for the rapid *in vitro* amplification of DNA sequences that flank a region of known sequence. The method uses the polymerase chain reaction (PCR), but it has the primers oriented in the reverse direction of the usual orientation. The template for the reverse primers is a restriction fragment that has been ligated upon itself to form a circle. This procedure of inverse PCR (IPCR) has many applications in molecular genetics, for example, the amplification and identification of sequences flanking transposable elements. In this paper we show the feasibility of IPCR by amplifying the sequences that flank an IS*1* element in the genome of a natural isolate of *Escherichia coli.*

THE polymerase chain reaction (PCR) is a powerful technique allowing the enzymatic amplification of specific regions of DNA without utilizing conventional cloning procedures. A major limitation of PCR as presently practiced is that it enables the amplification only of the region of DNA situated between two convergent primers. Any method for the *in vitro* amplification of the DNA sequences that flank a known segment of DNA would have many useful applications in genetics. Examples of applications include relatively easy identification of the consensus sequences for insertion of transposable elements and determination of the DNA sequences left behind after partial excision. In this paper, we demonstrate the feasibility of such a procedure using a combination of conventional techniques. The procedure is exemplified by the *in vitro* amplification of both upstream and downstream regions that surround an insertion sequence in a natural isolate of *Escherichia coli.* This application demonstrates a general approach using the inverse polymerase chain reaction (IPCR) for rapidly obtaining flanking regions of unknown sequences.

Typical PCR amplifications utilize oligonucleotide primers that hybridize to opposite strands. The primers are oriented such that extension proceeds inwards across the region between the two primers. Since the product of DNA synthesis of one primer serves as the template for the other primer, the PCR procedure of repeated cycles of DNA denaturation, annealing of primers, and extension by DNA polymerase results in an exponential increase in the number of copies of the region bounded by the primers (SAIKI *et al.* 1985; SCHARF, HORN and ERLICH 1986; FALOONA and MULLIS 1987). However, using the conventional PCR procedure, DNA sequences that lie immediately outside the primers are apparently inaccessible because oligonucleotides that prime DNA synthesis into flanking regions, rather than included regions, allow only a linear increase in the number of copies. The linear increase occurs because, for each primer, there is no priming of DNA synthesis in the reverse direction.

We used an extension of the polymerase chain reaction that permits the amplification of the regions that flank any DNA segment of known sequence, either upstream or downstream or both. In its effect, this conceptually simple technique, denoted "inverse PCR," allows one to "walk" outside a region of known sequence without resorting to conventional cloning procedures. The method has general applicability in genetics and can be used to establish rapidly the insertion sites of mobile genetic elements, to analyse the regions adjacent to a specific sequence, or to proceed along a stretch of uncharacterized DNA.

MATERIALS AND METHODS

The *Escherichia coli* strain studied was ECOR 64, which derives from the ECOR reference collection (OCHMAN and SELANDER 1984). Also designated strain C70, the strain was derived in Sweden from the urine of a woman with a urinary tract infection. For references about the strain and its classification by means of protein electrophoresis, see OCHMAN and SELANDER (1984).

Restriction digests were carried out using 5 μg of source DNA treated with 10 units of *Eco*RI according to the supplier's specifications (U.S. Biochemicals). Digested DNAs were electrophoresed through a 1.1% (w/v) agarose gel (SeaKem) in 1× TBE buffer (50 mM Tris, 100 mM Borate, 10 mM EDTA, pH 8.2). Appropriate fragments were excised from gel, electroeluted in 0.5× TBE, and extracted twice with phenol and once with chloroform; the DNA concentration was determined by UV spectrophotometry.

For circularization, 0.1 μg of the appropriate restriction fragment was diluted to a concentration of 0.5 μg/ml in ligation buffer (50 mM Tris HCl, pH 7.4, 10 mM $MgCl_2$, 10 mM dithiothreitol, 1 mM adenosine triphosphate and 10 μg/ml gelatin). The ligation reaction was initiated by the addition of T4 DNA ligase (New England Biolabs) to a concentration of 1 unit/μl and the reaction was allowed to proceed

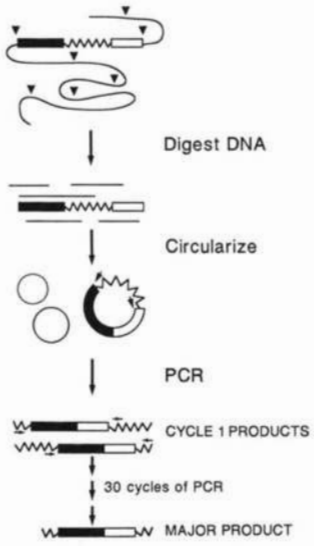

FIGURE 1.—Schematic of the inverse PCR procedure. The core region is depicted as a jagged line. Filled and open boxes represent upstream and downstream flanking regions, respectively. DNA is digested with a restriction enzyme (restriction sites denoted by triangles), circularized under conditions that favor the formation of monomeric circles, and enzymatically amplified using PCR. Oligonucleotide primers (constructed to anneal to the core region) and the direction of DNA synthesis are shown by arrows.

for 16 hr at 12°. The ligated sample was then treated with an equal volume of phenol:chloroform mixture, the aqueous phase was removed, and the DNA precipitated with ethanol and collected by centrifugation.

The PCR was performed manually in reactions containing 0.1 μg of circularized DNA obtained as described above in the presence of 50 pmol of each primer and 500 μM dNTPs. The primers were synthesized using an Applied Biosystems automated oligonucleotide synthesizer. We used 30 cycles of denaturation at 94° for 1.5 min, primer annealing at 48° for 1.0 min, and extension by *Taq* polymerase (Perkin-Elmer Cetus) at 70° for 4.0 min. The resulting sample was desalted and excess dNTPs were removed with a Centricon 30 microconcentration column from Amicon (HIGUCHI *et al.* 1988; SAIKI *et al.* 1988). The DNA products from the PCR reactions were fractionated a 1.1% agarose gel, denatured *in situ*, and transferred to an Amersham Hybond N nylon support membrane. The nylon filter was probed with a ^{32}P-labeled oligonucleotide primer using the methods in SAWYER *et al.* (1987).

RESULTS AND DISCUSSION

The basic method for obtaining the sequences flanking any particular region of interest (called the "core region") is shown schematically in Figure 1. Genomic DNA is cleaved using a restriction enzyme that has no restriction sites within the core region. Selection of appropriate restriction enzymes can be determined empirically by Southern blotting and hybridization procedures using all or part of the core region as probe. Selection of the appropriate fragment can be facilitated by computer search methods, since in many cases the entire nucleotide sequence of the core region will be known (*e.g.*, well characterized transposable elements or oncogenes). Optimally, the fragment of interest should be no greater than 2–3 kilobases (kb) longer than the core region, which is a limitation imposed by the size of a region that can be efficiently amplified using presently available methods of PCR. Future improvements in PCR methods

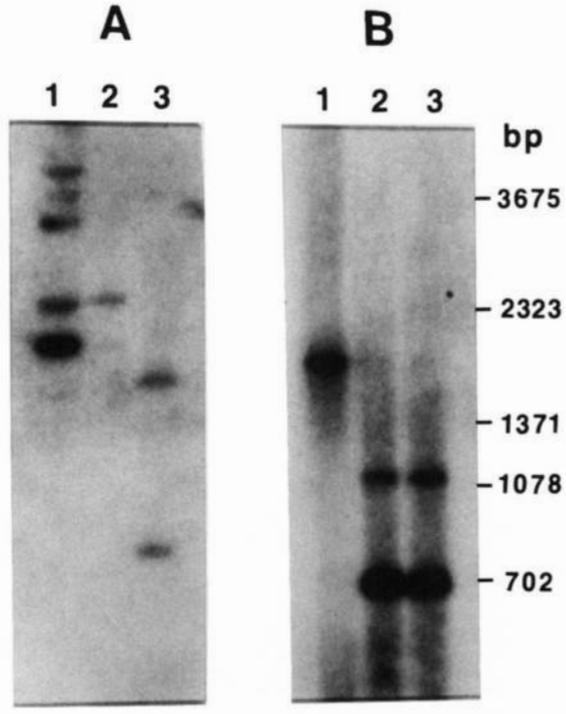

FIGURE 2.—Application of inverse PCR to amplify the regions flanking an IS*1* element in a natural isolate of *E. coli*. A, Genomic DNA from strain ECOR 64 was digested with *Eco*RI, electrophoresed in agarose, blotted and hybridized with a ^{32}P-labeled plasmid containing IS*1* (lane 1). To isolate a specific IS*1* and its flanking DNA, *Eco*RI fragments ranging from 2.2 to 2.5 kb were extracted from a preparative gel. This region contained a single IS*1* residing on a 2.4-kb *Eco*RI fragment (lane 2). When the purified *Eco*RI fragments containing a single IS*1* were digested with *Pst*I (which has a single site within IS*1*), two bands (corresponding to 0.8- and 1.65-kb fragments) hybridized with IS*1* (lane 3). B, Products of inverse PCR. DNA fragments containing a single IS*1* were circularized and subjected to 30 cycles of PCR using primers 305 and 306 (see Figure 3). Products of the reaction were electrophoresed in agarose, blotted, and hybridized to a ^{32}P-labeled oligonucleotide (primer 276), the sequence of which corresponds to the inverted repeat situated at the ends of IS*1* (see Figure 3). DNAs in lanes 1, 2 and 3 each contain 1% of the total amplification reaction. DNA in lane 1 is the intact product of PCR, that in lane 2 was digested with *Eco*RI, and that in lane 3 was digested with *Eco*RI and *Pst*I.

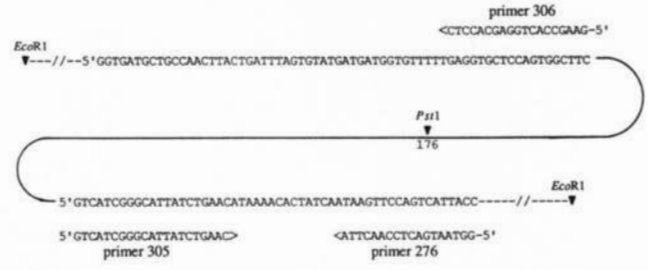

FIGURE 3.—Partial nucleotide sequence of IS*1* (OHTSUBO and OHTSUBO 1978) showing the location, composition and orientation of each oligonucleotide primer. The position of the single *Pst*I is also marked.

should allow much longer flanking sequences to be studied.

After restriction enzyme digestion, the DNA fragments produced by the restriction enzyme are diluted and ligated under conditions that favor the formation of monomeric circles (COLLINS and WEISSMAN 1984). The resulting intramolecular ligation products are then used as substrates for enzymatic amplification by PCR using oligonucleotide primers homologous to the ends of the core sequence, but facing in opposite orientations. The primary product of the resulting amplification is a linear double-stranded molecule including segments situated both 5′ and 3′ to the core region. The junction between the original upstream and downstream regions, otherwise ambiguous, can be identified as the restriction site of the restriction enzyme that was used to produce the linear fragments prior to ligation. By selecting a restriction enzyme that cleaves inside a known core sequence, the IPCR procedure will produce products containing only the upstream or only the downstream flanking regions.

To demonstrate the feasibility of IPCR, we used the method to amplify the DNA sequences that flank an IS*1* element in a natural isolate of *E. coli.* The insertion sequence IS*1* is 768 basepairs in length and, based on the published sequence (OHTSUBO and OHTSUBO 1978), it contains no sites for cleavage by *Eco*RI. For the feasibility demonstration we chose the *E. coli* strain ECOR 64 (OCHMAN and SELANDER 1984), which contains at least five copies of IS*1* (see lane 1 in Figure 2A, and SAWYER *et al.* 1987). DNA from ECOR 64 was digested with *Eco*RI and the restriction fragments were separated by electrophoresis in agarose. Fragments ranging 2.2 to 2.5 kb were excised from the gel and electroeluted in order to enrich for an IS*1* element that is located within a 2.45-kb *Eco*RI fragment (Figure 2A, lane 2).

Two oligonucleotide primers were synthesized, one complementary to nucleotides 45 through 63 (primer 305 in Figure 3) and the other identical to nucleotides 715 through 735 (primer 306) of IS*1*. These primers are situated near the 5′ and 3′ ends of IS*1*, but unlike the primers normally employed in PCR, they are oriented such that primer extension proceeds outward from the element (Figure 3). The products of the inverse PCR (lane 1 in Figure 2B) were subsequently treated with *Eco*RI, which yields two fragments corresponding to upstream and downstream flanking regions of this IS element (lane 2 in Figure 2B). The sizes of these fragments (0.7 and 1.1 kb) are in agreement with those predicted from Southern blots of genomic DNA. The *Eco*RI-digested, size fractionated DNA so obtained was also digested with *Pst*I, which is known to have a single cleavage site within IS*1* at nucleotide 171. The *Eco*RI-*Pst*I fragments from genomic DNA that hybridize with IS*1* are 0.8 and 1.65 kb in length (lane 3 in Figure 2A). As expected, these genomic fragments are approximately 0.10 and 0.55 kb longer than those produced by the *Eco*RI digestion of the inverse PCR product (lane 2 in Figure 2B). Moreover, the inverse PCR product contains no *Pst*I recognition site (lane 3 in Figure 2B), confirming that the amplified region does not include an intact IS*1* element.

The inverse PCR procedure permits the rapid amplification of regions of unknown sequence flanking a specified segment of DNA. Since, at present, only regions of limited size can be enzymatically amplified using PCR, and since primers must be synthesized from known sequences, the inverse PCR approach is not amenable to proceeding very long distances in the genome. However, the inverse PCR technique does eliminate the need to construct and screen genomic libraries in order to walk hundreds, if not thousands, of basepairs into flanking regions. In theory, inverse PCR could be carried out repeatedly allowing progressively more distant flanking sequences to be determined.

The authors would like to thank GINO CORTOPASSI, DAN GARZA and RUSS HIGUCHI for constructive aid during all aspects of the research. We are also grateful to ROBERT DUBOSE for providing unpublished information on the size distribution of restriction fragments containing IS*1* in natural isolates of *E. coli.*

LITERATURE CITED

COLLINS, F. S., and S. M. WEISSMAN, 1984 Directional cloning of DNA fragments at a large distance from an initial probe: a circularization method. Proc. Natl. Acad. Sci. USA **81:** 6812–6816.

FALOONA, F., and K. B. MULLIS, 1987 Specific synthesis of DNA in vitro via a polymerase chain reaction. Methods Enzymol. **155:** 335–350.

HIGUCHI, R. G., C. H. VON BEROLDINGEN, G. F. SENSABAUGH and H. A. ERLICH, 1988 DNA typing from single hairs. Nature **332:** 543–545.

OCHMAN, H., and R. K. SELANDER, 1984 Standard reference strains of *Escherichia coli* from natural populations. J. Bacteriol. **157:** 690–693.

OHTSUBO, H., and E. OHTSUBO, 1978 Nucleotide sequence of an insertion element, *IS1. Proc. Natl. Acad. Sci. USA* **75:** 615–619.

SAIKI, R., S. SHARF, F. FALOONA, K. B. MULLIS, G. T. HORN, H. A. ERLICH and N. ARNHEIM, 1985 Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science **230:** 1350–1354.

SAIKI, R. K., D. H. GELFAND, S. STOFFEL, S. J. SCHARF, R. G. HIGUCHI, G. T. HORN, K. B. MULLIS and H. A. ERLICH, 1988 Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239:** 487–491.

SAWYER, S. A., D. E. DYKHUIZEN, R. F. DUBOSE, L. GREEN, T. MUTANGADURA-MHLANGA, D. F. WOLCZYK and D. L. HARTL, 1987 Distribution and abundance of insertion sequences among natural isolates of *Escherichia coli.* Genetics **115:** 51–63.

SCHARF, S. J., G. T. HORN, and H. A. ERLICH, 1986 Direct cloning and sequence analysis of enzymatically amplified genomic sequences. Science **233:** 1076–1078.

Communicating editor: J. W. Drake