9-1-1989

Direct Genomic Sequencing of Bacterial DNA: The pyruvate kinase I gene of *Escherichia coli*

Osamu Ohara
Harvard University

Robert L. Dorit
Harvard University, rdorit@smith.edu

Walter Gilbert
Harvard University

Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 6883–6887, September 1989
Biochemistry

Direct genomic sequencing of bacterial DNA: The pyruvate kinase I gene of *Escherichia coli*

(polymerase chain reaction/expression/multiplex oligomer walking)

OSAMU OHARA*, ROBERT L. DORIT†, AND WALTER GILBERT

Department of Cellular and Developmental Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138

Contributed by Walter Gilbert, June 14, 1989

ABSTRACT **The genomic sequencing procedure is applied to the direct sequencing of uncharacterized regions of bacterial DNA by a "multiplex walking" approach. Samples of bulk *Escherichia coli* DNA are cut with various restriction enzymes, subjected to chemical sequencing degradations, run in a sequencing gel, and transferred to nylon membranes. When a labeled oligomer is hybridized to a membrane, a sequence ladder appears wherever the probe lies near a restriction cut. New probes, based on sequence that lies beyond other restriction sites, are then synthesized, and the membranes are reprobed to reveal new sequence. Repeated cycles of oligomer probe synthesis and subsequent reprobing permit rapid sequence walking along the genome. This oligomer walking technique was used to sequence the pyruvate kinase (EC 2.7.1.40) gene in *E. coli* without resorting to cloning or to library construction. The sequenced region was amplified by the polymerase chain reaction and subsequently transcribed and translated using both *in vivo* and *in vitro* systems, and the resultant gene product characterized to show that the gene encodes the type I isoform of pyruvate kinase.**

Genomic sequencing detects the DNA sequence of a particular molecule present in a heterogeneous mixture by exhibiting the sequence on a Southern blot, using a hybridizing labeled probe as an end label (1). The method examines the original genomic DNA molecules directly, without an intervening stage of amplification or cloning. It can detect sequence features in DNA from organisms with large, complex genomes and has been used to study the methylation patterns of animal (2–8) and plant (9) DNA. Footprinting and DNA interaction experiments in both prokaryotic and eukaryotic cells, *in vivo* (10–12) and *in vitro* (3, 13–15), have also taken advantage of this method. After a single set of chemical degradations and a single electrophoretic run, the DNA sequence lanes are transferred onto a membrane that can then be reprobed many times to display particular sequence ladders, making this a very fast sequencing method. Such an adaptation was devised to sequence inserts in both bacteriophage and plasmids by Richard Tizard and Harry Nick (personal communication, used in ref. 16). George Church has developed a very high-throughput, shotgun pattern of "multiplex" sequencing (17), which derives ≈50 times the usual sequence information from each gel and each set of chemistries. Here we wish to explore techniques to sequence bacterial DNA directly, moving step-by-step along the bacterial chromosome.

We hoped to extend the utility of genomic sequencing into previously unsequenced regions by using a "multiplex walking" strategy along the *Escherichia coli* chromosome. Aliquots of bulk genomic DNA from *E. coli* are digested with a variety of restriction enzymes, subjected to Maxam–Gilbert chemical sequencing reactions, run out in polyacrylamide sequencing gels, and finally transferred and crosslinked onto nylon filters. The filters thus contain adjacent tracks of restricted, chemically cleaved genomic DNA. When these filters are hybridized to an extremely radioactive probe, made by adding a tail of [^{32}P]AMP to a synthetic oligomer with terminal deoxynucleotidyltransferase, those lanes containing DNA that has been restricted at or near the hybridization site of the probe yield legible sequence.

This sequence information is then used to design new probes, derived from the 3′-most or 5′-most regions of the newly obtained sequence. The same nylon filters can then be reprobed with the new oligomers, providing additional new sequence. Repeated cycles of probe synthesis and reprobing of the nylon membranes enable one to walk briskly along the DNA in both 5′ and 3′ directions (Fig. 1).

To exemplify this oligomer walking method, we sequenced a pyruvate kinase (PK; EC 2.7.1.40) gene in *E. coli*. The comparison of all available PK sequences reveals regions of high amino acid sequence conservation (18–21). By synthesizing a long oligomer corresponding to one such region, we gained entry into the *E. coli* PK gene by carrying out genomic sequencing on an enriched genomic DNA mixture, using cross-hybridization with this heterologous probe. Repeated cycles of genomic sequencing and probe synthesis then provided the entire sequence of the gene.

Previous workers had suggested, on biochemical (22, 23) and on electrophoretic and genetic (24) grounds, that two noninterconvertible forms of PK are present in *E. coli*: type I (fructose 1,6-bisphosphate-activated) and type II (AMP-activated). Using the direct method, we could not *a priori* determine which of the two genes we had sequenced. Examination of the sequence could not resolve this question, because of an ambiguity created by the presence of two potential initiation sites. Thus, we amplified, transcribed, and translated the sequenced region to examine the gene product directly. On the basis of amino acid composition, molecular weight, and protein activity, we conclude that the gene we have sequenced encodes type I PK (PK-I).

MATERIALS AND METHODS

Materials. *E. coli* strains HB101, DH5α, and XL1-Blue were obtained from BRL and Stratagene. Terminal deoxynucleotidyltransferase was obtained from Boehringer Mannheim; *Thermus aquaticus* (*Taq*) polymerase was from Perkin–Elmer/Cetus; mung bean nuclease and exonuclease III were from BRL; all other DNA-modifying enzymes came from New England Biolabs. Buffers and conditions followed suppliers' specifications. Oligodeoxynucleotide primers and

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: PK, pyruvate kinase; PCR, polymerase chain reaction.

*Present address: Shionogi Research Laboratories, Fukushima-ku, Osaka 553, Japan.
†To whom reprint requests should be addressed.

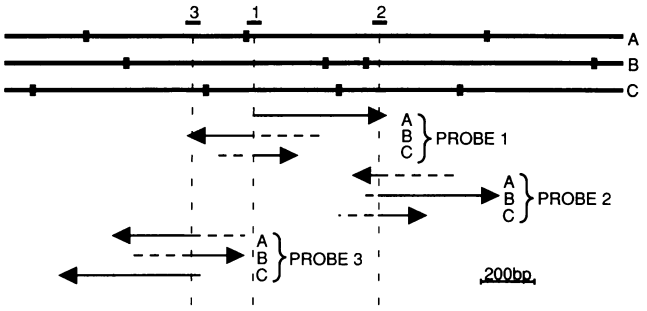

FIG. 1. The multiplex walking strategy. Total *E. coli* DNA is digested with an arsenal of 23 restriction enzymes (*Acc* I, *Alu* I, *Ava* II, *Apa*LI, *Ban* I, *Ban* II, *Bst*NI, *Cal* I, *Hae* II, *Hae* III, *Hin*cII, *Hin*dIII, *Kpn* I, *Mbo* II, *Pst* I, *Pvu* II, *Sau*3AI, *Nci* I, *Rsa* I, *Sty* I, *Taq* I, *Xho* I, *Xmn* I). Each restriction digest is subjected to a full set of Maxam–Gilbert sequencing chemistries, electrophoresed in denaturing polyacrylamide gels, electrotransferred to nylon filters, and immobilized by crosslinking to the filters. Each filter holds the chemical sequencing products of 8–10 different restriction digests. Filters are then probed with short, complementary, radioactively tailed oligonucleotides. Depending on the site of hybridization of the probe relative to a restriction cut, up to 300 base pairs (bp) of sequence can be directly visualized. In this diagram, the heavy lines represent three molecules of *E. coli* genomic DNA restricted with enzymes A, B, and C, respectively. The first cluster of three arrows shows the overall extent and orientation of sequence information that can be visualized from a single filter hybridized with probe 1. Each arrow shows the information obtained from each of the sequenced restriction digests (A, B, and C), given the indicated restriction sites. The solid portion of the arrow indicates legible sequence; the dashed portion originates at the relevant restriction cut. Sequence will be read in either the 5′ or 3′ direction, depending on the position of the probe relative to the restriction cuts. Sequence can be read along the length of a restriction fragment, beginning at the hybridization site and extending downstream to the position at which extraneous sequences, originating from the downstream restriction site, obscure the sequence ladder. The 3′-most and 5′-most sequence thus obtained is used to design probes 2 and 3, respectively. Sequence filters are subsequently stripped and reprobed with the newly designed oligomer probes, yielding additional sequence information as shown. This technique of filter probing, new probe design, and subsequent reprobing of the same nylon filters permits one to move rapidly in both directions along the length of a sequence of interest.

probes were synthesized on a Milligen 6500 synthesizer and were purified by electrophoresis in 20% polyacrylamide denaturing gels. Radiolabeled nucleotides, GeneScreen membranes, and the *in vitro* coupled transcription/translation system were purchased from NEN/DuPont and used according to their recommendations. Vector pBluescript KS(−) was obtained from Stratagene.

All computer analyses of DNA and amino acid sequences were carried out on a MicroVAX II, using the 1986 University of Wisconsin Genetics Computer Group (UWGCG) software package (41).

DNA Blot Analysis of Restriction Digests. Total *E. coli* genomic DNA was isolated from the HB101 strain (25) and subsequently cut with a variety of restriction enzymes (4- and 6-base-specific). Restriction digests of *E. coli* genomic DNA were run out in a 1% agarose gel containing 89 mM Tris/89 mM boric acid/2 mM EDTA (TBE buffer) and were subsequently transferred onto GeneScreen nylon membrane by standard capillary methods (26). After overnight transfer, DNA was crosslinked to the membrane by UV irradiation. Heterologous probes (e.g., C-1, Table 1) were radioactively labeled with [γ-^{32}P]dATP (5000–6000 Ci/mmol; 1 Ci = 37 GBq) by using T4 polynucleotide kinase. The membrane was probed at 45°C for 4 hr with 10 pmol of labeled oligonucleotide in 10 ml of genomic sequencing hybridization buffer [1% bovine serum albumin/0.5 M sodium phosphate (as Na^+ concentration), pH 7.2/7% $NaDodSO_4$/1 mM EDTA]. The membrane was washed three times (20 min per wash) at 45°C with 0.4 M NaCl/40 mM sodium phosphate, pH 7.2/1% $NaDodSO_4$/1 mM EDTA. Two final 10-min washes were carried out at 50°C using a solution containing 3.2 M tetramethylammonium chloride and 1% $NaDodSO_4$ (27). The wet nylon membrane was then wrapped in plastic wrap and exposed for 4–24 hr to Kodak X-Omat AR film.

Table 1. Genomic sequencing probes and polymerase chain reaction (PCR) primers

Name	Sequence (5′ to 3′)	Position*
	Oligonucleotides for genomic sequencing	
C-1	GCAGGGATCTCAATACCCAGGTCACCAC-GGGCCACCATAA	−1112
E-1	CGAGGATTTCGTCGAAGTTGTTG	−1058
E-2	TTTTCGATTTTGGAGATGTGGAT	−1023
E-3	CGCCTTGTTCGCAACCAAAGATC	−889
E-4	CTTCTGCGCGAGTCGGGCGTGGG	−1238
E-5	CAGTTTACGGTTGTCATTGTTG	−1411
E-6	GCCAACAGACAGGTCAGTAGTGA	−715
E-7	GCTCGCGCAGTACGTAAATAC	+1505
E-8	TAACATCTCTTCAGATTCGGT	−415
E-9	GATCTCTTTAACAAGCTGCGGCAC	−1624
E-10	CTCGCTCTAAGGATAGGTGAC	−260
E-11	ACGTCACCTTTGTGTGCCAGACCG	−1698
	Oligonucleotides for PCR	
E-12	GAGCTCTTCGATATACAAATTAATTC	−1802
E-13	AAGCTTGCGTAACCTTTTCCC	+5
C-2	CGTGGTGACCTGGGTATTGAGATCCC	+1085
C-3	GCGGTCTCCCCAGACAGCAT	−1302
C-4	ACCAAGGGACCTGAAATCCG	+554

*Relative to final PK sequence.

Fractionation and Enrichment of *Pst* I-Digested Target Fragment. To obtain legible sequence with cross-hybridizing probes, we enriched the DNA mixture for the PK target (see *Results*). DNA blot analysis showed that the 4.5- to 4.8-kilobase (kb) size range of a *Pst* I digest contained the entire PK sequence. *E. coli* DNA digested to completion with *Pst* I was run out in a 1% agarose gel, and fragments in the 4.5-kb range were isolated from the gel by transfer onto DEAE membrane (NA-45; Schleicher & Schuell) (28). The DNA in this size range (including the target) was eluted from the membrane by standard protocols, ethanol-precipitated, desalted, dried under vacuum, resuspended in water to a concentration of ≈20 ng/ml, and ligated into the dephosphorylated *Pst* I site of the pBluescript vector. The resultant plasmid mixture was amplified in *E. coli* strain DH5α and subsequently purified according to standard methods (29).

Genomic Sequencing. We sequenced DNA of interest by chemical degradation methods (30–32) using six reactions (G, A + G, A > C, C + T, C, and T). Approximately 5 μg of the appropriate DNA (plasmid mixture or total *E. coli* DNA) was used in each chemical sequencing reaction. Samples were then electrophoresed in either isocratic or gradient acrylamide denaturing sequencing gels (0.4 mm thick) (33), electrotransferred onto nylon membranes, and immobilized by UV crosslinking (1). Filters were probed with oligonucleotides that had been tailed with [α-^{32}P]dATP (6000 Ci/mmol) and subsequently purified on oligo(dT)-cellulose (34). Hybridizations were carried out as described above for restriction digests, except for a lower final probe concentration (0.2 nM) and the presence in the hybridization mixture of polyadenylic acid (20 μg/ml), added to prevent spurious probe binding due to the oligo(dA) tail. Sequencing filters were washed seven times (10 min per wash) at room temperature with a solution (preheated to 45°C) containing 0.2 M NaCl, 40 mM sodium phosphate (pH 7.2), 1% $NaDodSO_4$, and 1 mM EDTA. When imperfect probes were used, filters were also subjected to two final 3.2 M tetramethylammonium chloride

washes as previously described. Wet filters were wrapped in plastic wrap and exposed for 1–6 days to Kodak XAR-5 film without intensifying screens. Probes were stripped from the membranes by washing with 50 mM NaOH at room temperature for ≈20 min. Membranes were then rinsed with 50 mM Tris/EDTA buffer (pH 8.3), wrapped in plastic wrap, and stored at 4°C until the next reprobing.

Expression and Characterization of *E. coli* PK. After determining the DNA sequence of the PK gene, we designed and synthesized two flanking primers (E-12 and E-13, Table 1) to amplify a DNA fragment containing the entire PK gene by the PCR (35, 36). For ease in cloning, the primers were designed explicitly to contain restriction enzyme recognition sites at their 5′ ends (*Sac* I site for E-12; *Hin*dIII site for E-13). The PCR was carried out in a 100-μl volume containing 1 μg of *E. coli* total DNA, 100 pmol of each primer, 200 μM each dNTP, 2.5 units of *Taq* polymerase, 50 mM KCl, 10 mM Tris·HCl (pH 8.8), 10 μg of bovine serum albumin, 3 mM dithiothreitol, and 1.5 mM $MgCl_2$. Thirty cycles of amplification were carried out as follows: annealing, 50°C, 1 min; extension, 72°C, 2 min; denaturation, 94°C, 1 min; final extension, 72°C, 5 min. The amplified product of expected size (1.8 kb) was recovered from a 1% agarose gel by adsorption onto a DEAE membrane (28), subsequently eluted, and cloned into the *Pst* I site of the pBluescript vector by the "G-C tailing" method (37). The resultant construct is referred to as pEPK. Subsequently, we modified pEPK by deleting from the 5′ end of the insert with exonuclease III and mung bean nuclease (38). A new plasmid construct, determined by genomic sequencing to involve the deletion of 212 bp from the 5′ end of the PK gene, is referred to as pdEPK. Both of the plasmids were prepared using *E. coli* strain XL1-Blue, under standard procedures (29). *E. coli* cells harboring pBluescript, pEPK, or pdEPK were grown for 8 hr at 37°C in M9 medium supplemented with Casamino acids (0.2%) and ampicillin (100 μg/ml). The cells were ruptured by sonication and soluble protein fractions were prepared. We assayed PK activity by a modified lactate dehydrogenase coupled assay. The different regulatory properties of the two isozymes allowed us to measure the activity of type I and type II isozymes (PK-I and PK-II) separately (39). The soluble protein fractions were also analyzed by $NaDodSO_4$/PAGE (40). Protein concentrations were determined by Coomassie blue G250 binding (Bio-Rad protein assay kit).

The protein products encoded by the various plasmids were generated in the presence of [^{35}S]methionine by using an *in vitro* coupled transcription/translation system (NEN/DuPont) derived from *E. coli*. The products synthesized *in vitro* were analyzed by $NaDodSO_4$/PAGE and detected by autoradiography. Apparent molecular weights of the PK gene products were determined by comparison with protein molecular weight standards from Sigma and BRL.

RESULTS

Our intention was to sequence a gene directly on bacterial DNA by the genomic sequencing method. We initiated the sequencing of a PK gene in *E. coli* by using a heterologous probe derived from the chicken PK sequence. We first synthesized a 40-base-long oligomer (C-1) from a highly conserved region of amino acid sequence and tested whether this oligomer would hybridize specifically to a single region of the *E. coli* genome. A single band appeared in certain lanes when probe C-1 was hybridized to a DNA blot containing various restriction digests of total *E. coli* DNA (Fig. 2). This suggests that probe C-1 most likely cross-hybridizes to a bacterial PK sequence. However, when C-1 was used to probe membranes containing total *E. coli* DNA subjected to chemical sequencing reactions, using 5 μg of total DNA per

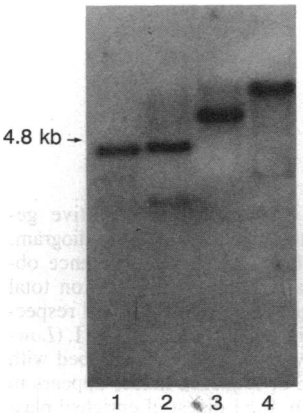

FIG. 2. Blot of digested total *E. coli* genomic DNA probed with heterologous oligomer C-1. Lanes: 1, *Pvu* II; 2, *Pst* I; 3, *Hin*dIII; 4, *Eco*RI. Autoradiogram was obtained by a 90-hr exposure with an intensifying screen, following a tetramethylammonium chloride wash.

sequencing reaction, a clear sequence ladder could not be discerned due to a poor signal-to-noise ratio.

We sought to improve this ratio by enriching for the signal. The DNA size range (4.5–4.8 kb) containing the target sequence was isolated from a *Pst* I digest of total *E. coli* DNA, inserted into a conventional vector, and then amplified in mass culture. The resultant amplified mixture of plasmids was digested with seven different restriction enzymes (*Apa*LI, *Hae* II, *Hae* III, *Hinc*II, *Pvu* II, *Stu* I, *Sty* I), since we did not know *a priori* which restriction sites would lie near the hybridization site of probe C-1. Using 5 μg of DNA (derived from the plasmid mixture) per sequencing reaction, we found that *Apa*LI-digested DNA produced a legible sequence ladder. This DNA sequence translated to a protein fragment similar to previously published eukaryotic PK sequences. Based on this initial fragment of *E. coli* PK sequence, we designed two perfect oligonucleotide probes (E-1 and E-2), which we then used for a new round of genomic sequencing (Fig. 3). Thus, by designing new probes as new sequence became available, and by reprobing previously used filters, we were able to move both upstream and downstream of our original entry point. The signal-to-noise problem encountered with the heterologous C-1 probe disappeared when we used perfect probes. In this manner, we were able to read sequences of 200–400 bp with overnight to 5-day exposures. The synthesis of new probes and the labeling and reprobing took as little as 3 days per cycle. Fig. 4 shows a representative fragment of the PK sequence, visualized directly either from total *E. coli* DNA or from an enriched mixture by using perfect probes.

Using this walking strategy for genomic sequencing, we determined a 1.8-kb nucleotide sequence that included a long open reading frame. The predicted amino acid sequence shows extensive similarity to all previously sequenced PKs. However, two important issues could not be resolved on the

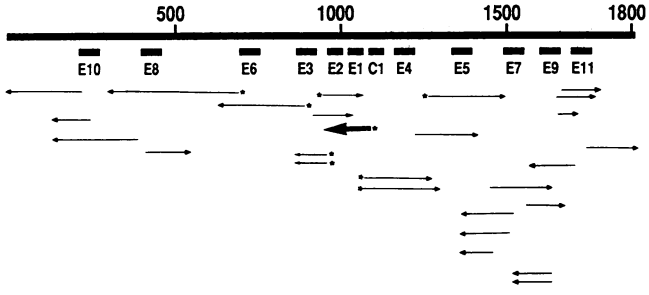

FIG. 3. Sequencing strategy, showing the extent of sequence information obtained using each of the listed oligomer probes. Positions of the oligomers are shown relative to the final numbering of the PK sequence (oligomers not to scale). Asterisks indicate sequences obtained from enriched plasmid mixture; others were obtained from total genomic DNA. Heavy arrow marks point of entry into sequence with heterologous probe.

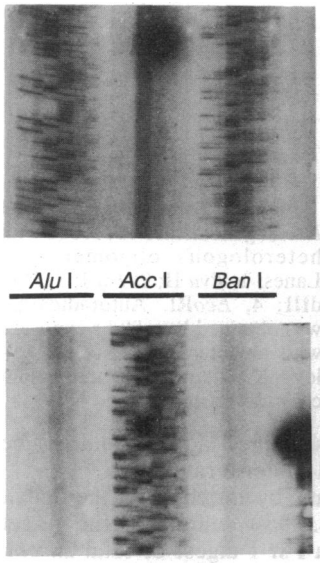

FIG. 4. Representative genomic sequence autoradiogram. (*Upper*) Detail of sequence obtained using probe E-3 on total genomic DNA restricted respectively with *Alu* I and *Ban* I. (*Lower*) The same filter, probed with E-8; sequence ladder appears in an *Acc* I digest of enriched plasmid mixture. Lanes show products of Maxam–Gilbert reactions, in following order: G, A + G, A > C, T + C, C, and T. Six-day exposures were made without intensifying screens.

basis of the DNA sequence alone: (*i*) the actual initiation codon, since two possible candidates for the initiating methionine are present (bases 140 and 355), and (*ii*) the identity of the PK isotype (I or II).

The two isotypes of PK in *E. coli* have been reported to have molecular weights of 56,000 and 51,000, respectively. Their overall amino acid compositions have been published and their sizes estimated at 522 residues (type I) and 488 residues (type II) (23). The *E. coli* sequence we obtained predicts possible proteins of 534 and 462 amino acids, while the amino acid composition suggests that the sequenced gene corresponds to type I.

To resolve the identity of this gene, we examined the gene product. Fig. 5 shows two expression plasmids constructed for the purpose: pEPK contains the full (1.8-kb) PCR-amplified PK gene, inserted in pBluescript in the opposite direction to the *lac* promoter; pdEPK is a truncated derivative. To determine the site of translation initiation, we attempted to delete between 140 and 355 bp from pEPK. After analyzing the deletion products, we selected one, pdEPK, from which nucleotides 1–212 had been deleted, thus removing the first methionine codon at base 140.

When plasmids pEPK and pdEPK were translated in a cell-free transcription/translation system derived from *E. coli*, two polypeptides (M_r 56,000 and 54,000) appeared on $NaDodSO_4$/PAGE, along with the ampicillin-resistance gene product (Fig. 6*A*), although pdEPK produced smaller amounts of both polypeptides. Further, when extracts of *E. coli* cells transformed by these two plasmids were run out on $NaDodSO_4$/PAGE, they consistently showed a strong doublet that comigrated with the M_r 56,000/54,000 *in vitro* products of pEPK (Fig. 6*B*). Since the deletion plasmid and the intact construct produce polypeptides of the same length,

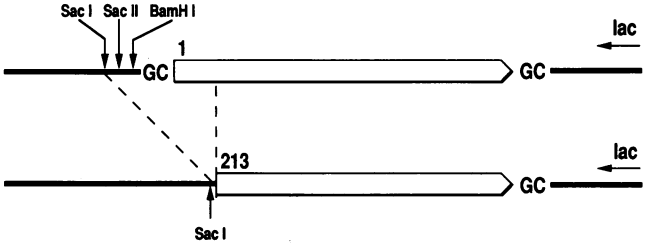

FIG. 5. pEPK and pdEPK constructs, derived from pBluescript vector. Open bars indicate the PCR-amplified PK gene either ligated directly into pBluescript by homopolymer tailing or after partial deletion at the 5′ end. Orientation of *lac* promoter is indicated.

the methionine codon at position 140 can be ruled out as the initiation codon.

Table 2 shows the PK enzymatic activity in crude extracts of *E. coli* cells harboring pEPK, pdEPK, or pBluescript. Since PK-I has only about 1% of its activity in the absence of fructose 1,6-bisphosphate, we could separately estimate PK-I and PK-II activity in the crude cell extracts. Both plasmids significantly stimulated PK-I activity (200-fold and 20-fold, respectively). We attribute the apparent increase in PK-II activities to the residual PK-I activity present at the 1% level in the PK-II fraction. We conclude that pEPK contains both the promoter and coding sequence of the *E. coli* PK-I gene, while the 212-bp deletion at the 5′ end of the insert in pdEPK eliminates much of the PK-I promoter activity but does not cut into the coding sequence.

We were surprised not to see any evidence by cross-hybridization for a PK-II gene. We then sought to find such a gene by PCR amplification of *E. coli* DNA, using two primers constructed from conserved regions of chicken PK. However, that amplification yielded only one major band, which, on sequencing, turned out to be the PK-I gene. We cut that band with two restriction enzymes (*Apa* I, *Bcl* I) to see whether any PK-II DNA fragments might be comigrating with the PK-I fragment; only PK-I sequences were detected.

DISCUSSION

These results extend the utility of genomic sequencing by demonstrating that it can be used, with a multiplex walking strategy, to decipher bacterial DNA sequences. Repeated cycles of sequencing, probe synthesis, and hybridization allowed us to advance rapidly both upstream and downstream from our original point of entry onto the *E. coli* genome and yielded the sequence of the entire PK-I gene.

However, we were not able to obtain readable sequence patterns from bacterial DNA by using imperfect, cross-hybridizing primers. We solved this problem by size-selecting particular fragment lengths (in this case, 4.5-kb *Pst* I fragments containing the entire PK gene) and thus reducing the background. Alternatively, we could have used imperfect primers in conjunction with the PCR. In fact, two PCR primers derived from conserved regions of the chicken PK sequence were successfully used to amplify a region of the *E.*

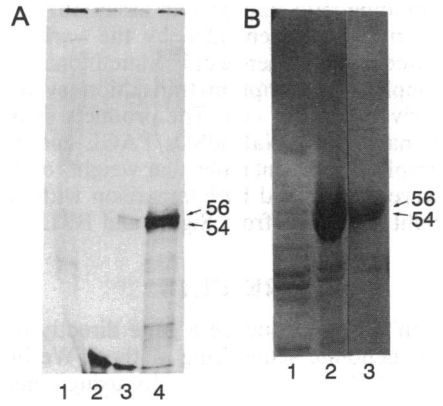

FIG. 6. (*A*) Autoradiogram of *in vitro* translation products labeled with [^{35}S]methionine. Lanes: 1, control, no DNA added; 2, control, intact pBluescript vector; 3, deletion plasmid pdEPK products; 4, plasmid pEPK products. (*B*) Autoradiogram of *in vivo* translation products obtained by using the soluble fraction of *E. coli* cell extracts. Total protein loaded, determined by Coomassie blue G250 binding assay, is indicated in parentheses. Lanes: 1, control, intact pBluescript vector (100 μg); 2, pEPK (150 μg); 3, pEPK (40 μg). Products were run in 8% $NaDodSO_4$/polyacrylamide gels and stained with Coomassie brilliant blue R250. The M_r 56,000 and 54,000 bands are indicated.

Table 2. PK-I and PK-II activities of plasmid constructs

Plasmid	Activity, units/mg of protein		
	PK-I	PK-II	Adjusted PK-II
pBluescript	0.24 (1.0)	0.04 (1.0)	—
pEPK	>50 (>208)	0.74 (18.5)	0.04
pdEPK	5.0 (20.8)	0.12 (3.0)	0.05

Numbers in parentheses indicate activities relative to pBluescript control. Adjusted PK-II activity: observed PK-II activity − residual PK-I activity (assumed to be 1.4% of total activity). Unit activity is proportional to amount of enzyme present.

coli PK gene for sequencing. Heterologous PCR, when successful, may represent the fastest route of entry into new sequences. Once a novel sequence is entered, either by using imperfect probes on enriched target DNA or by PCR, genomic sequencing provides a rapid, general method for brisk walking along sequences of interest. A single set of nylon membranes, containing genomic DNA digested with a battery of restriction enzymes and subjected to Maxam–Gilbert chemistries, can be probed, stripped, and reprobed up to 40 times with newly synthesized oligomers, yielding sequence in both directions from the point of entry. By obviating the need for library construction and screening and by reusing the products—already fixed onto a nylon filter—of a set of sequencing chemistries, our methods facilitate and speed the sequencing process and eliminate the risk of artifacts generated during library construction or subcloning.

Although this multiplex oligomer walking technique could be applied to genomes of much higher complexity, the practical limitation is the time involved in the exposures needed to visualize the sequence. Methods of labeling probes to higher specific activities or more sensitive detection techniques would eliminate the bottleneck.

The results of sequencing can be coupled, via PCR, to the analysis of gene products. We inserted PCR-generated sequences of the complete PK gene into plasmids that could be expressed both *in vivo* and *in vitro*. The PCR product also can function directly in an *in vitro* coupled transcription/translation system to yield the same protein products as the plasmid construct.

Translation, both *in vivo* and *in vitro*, of the pEPK plasmid consistently produced two bands on NaDodSO$_4$/PAGE; *in vitro*, the PK-I sequence directs the synthesis of two polypeptides of M_r 56,000 and 54,000. Posttranslational modification of the protein is not likely in the cell-free system, so these two distinct products probably arise by an inappropriate initiation of translation from an internal methionine codon. The smaller band appears too large to be the PK-II product and does not show any PK-II activity. Thus PK-I is 462 amino acids long yet has an apparent molecular weight of 56,000, in contradiction to the length (522 amino acids) predicted by previous authors. We attribute this discrepancy to the anomalous migration of PK-I in NaDodSO$_4$/PAGE. The amino acid composition deduced from our PK-I sequence is in close accord with that reported by other workers, although we do find a single tryptophan residue not identified previously. It is unclear whether the two gene products appearing in the doublet are made by *E. coli* under normal circumstances. The second gene product has not been observed before, although if it does not exhibit PK-I activity, it is unlikely to have been noticed previously.

Table 3. PK amino acid sequence comparisons

	E. coli	Yeast	Chicken
Yeast (*Saccharomyces cerevisiae*)	43%		
Chicken (*Gallus gallus*)	47%	47%	
Rat (*Rattus norvegicus*)	44%	48%	68%

Values indicate the percentage of identical amino acids shared in a given pairwise comparison. Sequences were aligned using the algorithm of Needleman and Wunsch as implemented in the UWGCG software package (41).

PK-I is homologous to the PKs of chicken, rat, cat, and yeast. Table 3 shows the extent of amino acid sequence conservation. However, we could not detect the PK-II gene by cross-hybridization or PCR, suggesting that it is highly divergent from, or perhaps unrelated to, the PK-I gene.

Our results show that the power and relative ease of genomic sequencing can be used to decipher unknown sequences directly from bacterial DNA, providing a rapid, general method for the characterization of genes.

We thank Milligen for the use of their 6500 synthesizer, Tom Maniatis for critical reading of the manuscript, members of the Gilbert lab for helpful discussions, and L. Schoenbach for the illustrations. This work was supported by the National Institutes of Health (Grant GM37997-02). O.O. was supported by Shionogi Research Laboratories.

1. Church, G. M. & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1991–1995.
2. Ephrussi, A., Church, G. M., Tonegawa, S. & Gilbert, W. (1985) *Science* **227**, 134–140.
3. Church, G. M., Ephrussi, A., Gilbert, W. & Tonegawa, S. (1985) *Nature (London)* **313**, 798–801.
4. Nick, H. & Gilbert, W. (1985) *Nature (London)* **313**, 795–798.
5. Saluz, H. P., Jiricny, J. & Jost, J. P. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 7167–7171.
6. Saluz, H. & Jost, J. P. (1986) *Gene* **42**, 151–157.
7. Becker, P. B., Ruppert, S. & Schutz, G. (1987) *Cell* **51**, 435–443.
8. Saluz, H. P., Feavers, I. M., Jiricny, J. & Jost, J. P. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6697–6700.
9. Nick, H., Bowen, B., Perl, R. J. & Gilbert, W. (1986) *Nature (London)* **319**, 243–246.
10. Selleck, S. B. & Majors, J. (1987) *Nature (London)* **325**, 173–177.
11. Pauli, V., Chrysogelos, S., Stein, G., Stein, J. & Nick, H. (1987) *Science* **236**, 1308–1311.
12. Bushman, F. P. & Ptashne, M. (1988) *Cell* **54**, 191–197.
13. Richet, E., Abcarian, P. & Nash, H. (1986) *Cell* **46**, 1011–1021.
14. Fujiwara, T. & Mizuuchi, K. (1988) *Cell* **54**, 497–504.
15. Hromas, R., Pauli, U., Marcuzzi, A., Lafrenz, D., Nick, H., Stein, J., Stein, G. & Van-Ness, B. (1988) *Nucleic Acids Res.* **16**, 953–967.
16. Marchionni, M. & Gilbert, W. (1986) *Cell* **46**, 133–141.
17. Church, G. M. & Kieffer-Higgins, S. (1988) *Science* **240**, 185–188.
18. Lonberg, N. & Gilbert, W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3661–3665.
19. Burke, R. L., Tekamp-Olson, P. & Najarian, R. (1983) *J. Biol. Chem.* **258**, 2193–2201.
20. Lone, Y. C., Simon, M. P., Kahn, A. & Marig, J. (1986) *FEBS Lett.* **195**, 97–100.
21. Muirhead, H., Clayden, D. A., Barford, D., Lorimer, C. G., Fothergill-Gilmore, L. A., Schiltz, E. & Schmitt, W. (1986) *EMBO J.* **5**, 475–481.
22. Somani, B. L., Valentini, G. & Malcovati, M. (1977) *Biochim. Biophys. Acta* **482**, 52–63.
23. Valentini, G., Iadarola, P., Somani, B. L. & Malcovati, M. (1979) *Biochim. Biophys. Acta* **570**, 248–258.
24. Garrido, P. A. & Cooper, R. A. (1983) *FEBS Lett.* **162**, 420–422.
25. Sato, H. & Miure, K. I. (1963) *Biochim. Biophys. Acta* **72**, 619–629.
26. Southern, E. (1975) *J. Mol. Biol.* **98**, 503–517.
27. Wood, W. I., Gitschler, J., Lasky, L. A. & Lawn, R. M. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1585–1588.
28. Dretzen, G., Bellard, M., Sassone-Corsi, P. & Chambon, P. (1981) *Anal. Biochem.* **112**, 295–298.
29. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
30. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
31. Rubin, C. M. & Schmid, C. W. (1980) *Nucleic Acids Res.* **8**, 4613–4619.
32. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
33. Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3963–3965.
34. Collins, M. L. & Hunsker, W. R. (1985) *Anal. Biochem.* **151**, 211–224.
35. Scharf, S. J., Horn, G. T. & Erlich, H. A. (1986) *Science* **233**, 1076–1078.
36. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988) *Science* **239**, 487–491.
37. Eschenfeldt, W. H., Puskas, R. S. & Berger, S. L. (1987) *Methods Enzymol.* **152**, 337–342.
38. Henikoff, S. (1984) *Gene* **28**, 351–359.
39. Malcovati, M. & Valentini, G. (1982) *Methods Enzymol.* **90**, 170–179.
40. Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685.
41. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.