Genetics. In the article “Use of the DNA polymerase chain reaction for homology probing: Isolation of partial cDNA or genomic clones encoding the iron–sulfur protein of succinate dehydrogenase from several species” by Stephen J. Gould, Suresh Subramani, and Immo E. Scheffler, which appeared in number 6, March 1989, of Proc. Natl. Acad. Sci. USA (86, 1934–1938) the partial amino acid sequence reported for the iron–sulfur protein of succinate dehydrogenase from Drosophila melanogaster is in error. The entire gene has recently been cloned (H. Au and I.E.S., unpublished work), using the same partial cDNA clone in the screening of a D. melanogaster genomic library. There are a significant number of sequence changes (approximately 72% sequence identity with the mammalian coding sequence), which make us believe that the original clone arose from a mix-up or a contamination. The paper described a relatively unusual, at the time, application of the polymerase chain reaction, and it was illustrated by cloning partial genomic or cDNA sequences from several species. The human and Saccharomyces cerevisiae sequences have since been confirmed independently by us and by others (1, 2).


Biochemistry. In the article “Marek disease virus encodes a basic-leucine zipper gene resembling the fos/jun oncogenes that is highly expressed in lymphoblastoid tumors” by Dan Jones, Lucy Lee, Juinn-Lin Liu, Hsing-Jien Kung, and Joanne K. Tillotson, which appeared in number 9, May 1, 1992, of Proc. Natl. Acad. Sci. USA (89, 4042–4046), the authors request that the following correction be noted. In the sequence presented in Fig. 2 on p. 4044, a G residue should be inserted after base +1053. This sequence correction does not affect the protein motifs (basic/leucine zipper or proline repeats) reported in the earlier paper. The correct DNA sequence has been deposited in GenBank.

Cell Biology. In the article “Interaction with basement membrane serves to rapidly distinguish growth and differentiation pattern of normal and malignant human breast epithelial cells” by Ole William Petersen, Lone Rønnov-Jessen, Anthony R. Howlett, and Mina Bissell, which appeared in number 19, October 1, 1992, of Proc. Natl. Acad. Sci. USA (89, 9064–9068), the authors request that the following correction be noted. On p. 9065, left column, line 1, μg/ml should be replaced with mg/ml.

Neurobiology. In the article “Physiological and anatomical evidence for a magnocellular defect in developmental dyslexia” by Margaret S. Livingstone, Glenn D. Rosen, Frank W. Drislane, and Albert M. Galaburda, which appeared in number 18, September 15, 1991, of Proc. Natl. Acad. Sci. USA (88, 7943–7947), the authors request that the following correction be noted. On p. 7944, the first sentence of the first paragraph should read as follows: “We examined the LGN in autopsies specimens from five dyslexic subjects (four males and one female; mean age, 34.2 ± 13.7 years) and five nondyslexic subjects (four males and one female; mean age, 40 ± 11.2 years).”

Biochemistry. In the article “Purification and characterization of cytosolic aconitate from beef liver and its relationship to the iron-responsive element binding protein” by Mary Claire Kennedy, Liane Mende-Mueller, George A. Blondin, and Helmut Beinert, which appeared in number 24, December 15, 1992, of Proc. Natl. Acad. Sci. USA (89, 11730–11734), the authors wish that the following correction be noted. In Table 4, entry 6 (amino acid sequence of domain 4), Arg’ should be below the first Lys (position 732) and not below the second Lys (position 736).
Use of the DNA polymerase chain reaction for homology probing: Isolation of partial cDNA or genomic clones encoding the iron–sulfur protein of succinate dehydrogenase from several species

STEPHEN J. GOULD, SURESH SUBRAMANI, AND IMMO E. SCHEFFLER

Department of Biology and Center for Molecular Genetics, University of California, San Diego, La Jolla, CA 92093

Communicated by E. Peter Geiduschek, December 12, 1988 (received for review November 9, 1988)

ABSTRACT The DNA polymerase chain reaction was developed for in vitro amplification of specific DNA sequences, and it has been used for a wide variety of purposes in several fields. We have developed an application of the polymerase chain reaction that is useful for the isolation of partial cDNA or genomic clones of conserved genes. We used this technique to clone the gene encoding the iron protein subunit (27 kDa) of succinate dehydrogenase (EC 1.3.5.1) from several species, including human, rat, Drosophila melanogaster, Arabidopsis thaliana, Schizosaccharomyces pombe, and Saccharomyces cerevisiae. Mixed oligonucleotide primers corresponding to two conserved regions of the protein were used in conjunction with genomic and cDNA templates in the reaction. The primers contained all possible nucleotide combinations that could encode the corresponding peptide sequences. These oligonucleotide mixtures contained 262,144 (216) and 8192 (213) unique sequences, respectively. Use of the polymerase chain reaction for homology probing allows one to utilize more complex mixtures of oligonucleotides as probes than is possible with filter hybridization screening techniques. In addition, the polymerase chain reaction offers the advantage of synthesizing the DNA product directly, in some cases obviating the need to construct cDNA or genomic libraries. This application of the polymerase chain reaction should be useful not only for the identification of conserved genes in a variety of species but also for the isolation of previously unknown members of gene families.

Many methods have been described that facilitate the isolation of desired genes. However, most require the use of reagents that may be difficult to obtain, such as antibodies specific for the protein encoded by the gene (for antibody screening of cDNA expression libraries) or a homologous or related gene from a different organism (for reduced-stringency hybridization-screening of either genomic or cDNA libraries). In instances where the only information available is the amino acid sequence known or presumed to be present in the protein, homology probing has been used to identify the gene of interest (1, 2). Homology probing involves screening cDNA or genomic libraries with radioactively labeled synthetic oligonucleotide probes corresponding to amino acid sequences of the protein. Unfortunately, this technique has certain limitations, one of which is the rather severe restriction on the degeneracy of the oligonucleotides that can be used (3).

To overcome some of the limitations of the oligonucleotide-hybridization approach to gene cloning, we have developed a variation of homology probing that involves the DNA polymerase chain reaction (4–6) and allows the use of highly degenerate oligonucleotides as probes. This method requires only knowledge of the amino acid sequence of two short regions (approximately 7–10 amino acids in length) of the encoded protein. Two degenerate oligonucleotide mixtures corresponding to each known peptide sequence are used as primers in the reaction. The primers are fully degenerate and contain all possible codon combinations that could encode the known amino acid sequences. The template for the amplification may be any of several DNA sources, including genomic DNAs and supercoiled forms of plasmid libraries.

We have used this modified version of homology probing to isolate partial cDNA and genomic clones of the gene encoding the small subunit of succinate dehydrogenase (SDH; EC 1.3.5.1). SDH is an essential enzyme of the tricarboxylic acid cycle and is also a component of complex II of the mitochondrial electron transport chain (7, 8). Thus it is indispensable for respiration. The enzyme catalyzes the transfer of electrons from succinate to ubiquinone, releasing fumarate. SDH contains two subunits, a 70-kDa ferroflavoprotein (FP) and a 27-kDa iron protein (IP) (8). The existence of a [4Fe–4S] cluster in the 27-kDa subunit appears to be established, on the basis of physical measurements and protein sequence data (9, 10). In a current model the IP subunit forms the bridge between the large FP subunit and the integral membrane proteins that make up the rest of complex II (10). Because of its relatively simple composition, complex II and its components represent ideal models with which to elucidate the mechanisms involved in the assembly of multisubunit complexes within the mitochondria.

The amino acid sequence of the bovine SDH IP subunit has been compared with the deduced amino acid sequence of the analogous Escherichia coli protein (11–13). Two highly degenerate oligonucleotide mixtures (262,144 and 8192 unique sequences, respectively) corresponding to two peptides conserved between the bovine and E. coli SDH IP subunits were used as primers in the polymerase chain reaction, and the templates were either genomic DNAs or plasmid forms of genomic or cDNA libraries from species as diverse as yeasts and mammals. The reaction products derived from each DNA template were cloned and confirmed to be IP subunit genes by sequence analysis. A search of the PIR protein sequence data bank (version 56.0) confirmed the existence of homologs to other Fe–S proteins such as the ferredoxins.

This report demonstrates the usefulness of combining the polymerase chain reaction with homology probing for the isolation of a gene from multiple species. Studies of evolutionary relationships between organisms, rates of sequence divergence during evolution, and comparisons of functionally important domains of proteins will be greatly facilitated by this technique.

MATERIALS AND METHODS

Reagents. The human primary fibroblast, mouse C3H10T½, and rat cerebral cortex cDNA libraries were obtained from H. Okayama (National Institutes of Health), the Xenopus laevis

Abbreviations: SDH, succinate dehydrogenase; IP, iron protein.
cDNA library was from W. Dunphy (Department of Biology, University of California, San Diego), the Drosophila melanogaster cDNA library was from N. Brown (The Biological Laboratories, Harvard University), the Schizosaccharomyces pombe genomic library was from L. Clarke (Department of Biology, University of California, Santa Barbara), the Saccharomyces cerevisiae genomic library was from M. Yaffe (Department of Biology, University of California, San Diego), the Arabidopsis thaliana genomic DNA was from N. Crawford (Department of Biology, University of California, San Diego), and the Dicyostelium discoideum genomic DNA was from J. Dynes (Department of Biology, University of California, San Diego). Oligonucleotides were obtained from Operon Technologies (San Pablo, CA). Thermus aquaticus (Tag) DNA polymerase was purchased from Perkin-Elmer/Cetus. All other reagents were obtained from standard sources.

Polymerase Chain Reaction Conditions, Cloning, and Sequencing. The conditions for the polymerase chain reactions were as follows: Ten micrograms of each set of primers (A and B), 100 ng of template DNA (either supercoiled DNA amplified from purified cDNA libraries or phage DNA from an AGT11 cDNA library, or purified genomic DNA) were combined with 10 μl of 10× Taq polymerase buffer (50mM KCl/100 mM Tris-HCl, pH 8.3/15 mM MgCl2/0.1% gel) and 16 μl of a dNTP solution (1.25 mM dGTP/1.25 mM dCTP/1.25 mM dATP/1.25 mM dTTP). The volume was brought to 100 μl with H2O and the mixture was overlaid with 100 μl of mineral oil to prevent evaporation. The tube was heated to 94°C for 2 min to denature the template. Then 2.5 units of Taq DNA polymerase was added and the solution was heated at 94°C for 2 min. At this point the cycling reaction was initiated. Denaturation was for 1.5 min at 94°C, annealing for 2 min at 50°C, 55°C, 60°C, or 65°C, and polymerization for 3 min at 72°C. The cycle was repeated 30 times. The product was digested with EcoRI and Sac I (the recognition sites for these enzymes were incorporated at the 5′ ends of the A and B primers, respectively) and size fractionated by agarose gel electrophoresis. In each case a DNA molecule of the predicted size was isolated, ligated with EcoRI- and Sac I-digested pUC19, and used to transform E. coli to ampicillin resistance. Independent clones containing inserts of the proper size were isolated and the DNA sequences of the inserts were determined by using the dideoxy chain-termination method with modified T7 DNA polymerase (Sequenase, United States Biochemical).

RESULTS

Synthesis of a Partial Human SDH IP cDNA. The primary structures of both the E. coli and the bovine SDH IP proteins

PEPTIDE A

\[
\text{Cys Arg Glu Gly Ile Cys Gly Ser Cys Ala Met}
\]

PRIMER A

5′-GAA TTC TGG GSN GAR GGN ATN TGY GGN ZNN TGY GGN ATG-3′

(EcoRI)

Degeneracy = \(2^{18}\) or 262144

PEPTIDE B

\[
\text{Cys His Thr Ile Met Asn Cys Thr Glu Thr Cys}
\]

PRIMER B

3′-ACR GTR TGN TAN TAC TTR ACR TGN CTG TGN ACR CTC GAG-5′

(SacI)

Degeneracy = \(2^{13}\) or 8192

Fig. 1. Sequences of the two conserved peptides of the SDH IP and corresponding degenerate oligonucleotides used as primers. Peptide A spans amino acids 65–75 in the bovine SDH IP protein sequence, while peptide B represents amino acids 215–225 of this protein. Each amino acid sequence was backtranslated to all possible DNA sequences that could potentially encode the stretch of amino acids. The oligonucleotide sequences of primers A and B are presented below the amino acid sequences of the respective peptides. At the 5′ end of each primer, additional sequences were appended to the primers to allow restriction digests of the polymerase chain reaction product with EcoRI (primer A) and Sac I (primer B) for subsequent cloning and sequence analysis. The nucleotide sequences are presented in the following code: G, deoxyguanosine; C, deoxycytidine; A, deoxadenosine; T, deoxythymidine; Y, either T or C; R, either A or G; N, A, G, C, or T; S, either C or A; Z, either T or A; X, either C or G.
We have used the polymerase chain reaction for in vitro amplification of partial cDNA and genomic clones of the SDH IP subunit from human, rat, D. melanogaster, A. thaliana, S. pombe, and S. cerevisiae. Subsequent experiments have shown that the same conditions can be used to isolate partial cDNA or genomic clones of the SDH IP gene from mouse, Xenopus laevis, and Dictyostelium discoideum. In these reactions we were able to use highly degenerate oligonucleotide primers (>10^4 different sequences) as well as a variety of template DNAs (phage (AGT11 cDNA library in the case of X. laevis), genomic, and supercoiled plasmid DNAs). In all cases, a DNA molecule with the molecular weight predicted for amplification of the SDH small subunit gene was observed after agarose gel electrophoresis of the reaction product. It should be emphasized that the DNA molecule of predicted size was the predominant, if not the only, double-stranded product on the gel. We also observed that the desired product appeared consistently when the annealing step was at 50–60°C, but not at 65°C. The product from the most stringent, successful condition was chosen in each case.

Clones derived from each amplification reaction containing inserts of the proper size were isolated at a high frequency (approximately 80%) during the initial cloning step. Sequence analysis of two clones isolated from each template source demonstrated that all (12/12) encoded the SDH IP from one or another species.

The experiments we have presented in this paper demonstrate the usefulness of combining homology probing with the polymerase chain reaction. Until this report, homology probing has involved screening cDNA or genomic libraries (in the form of plaque or colony lifts) with radiolabeled oligonucleotide probes. This approach has been successful for the isolation of known genes from several organisms (14, 15) and the identification of novel members of gene families [e.g., protein kinases (1), and guanine nucleotide-binding (G) proteins (2, 3)]. However, there are certain limitations to this technique. In the conventional use (filter hybridization) of degenerate oligomers as probes in the screening of libraries, the strength of a signal from a positive clone will depend on the specific activity of the oligonucleotide (cpm per nucleotide phosphate) and on the actual amount of the oligonucleotide that is bound stably to the target DNA. At degeneracies greater than 500–1000, and assuming realistic specific activities of the probes, the actual amount of the relevant oligonucleotide may be too small to provide a useful signal-to-noise ratio [previously discussed by Touchot et al. (3)]. To compensate for this limitation one must take into account codon usage information in the design of the probes and make use of only a subset of the potential codons. In doing so, one introduces a bias against the detection of genes that utilize unusual codons and also against the detection of genes in organisms with a different codon usage.

In contrast, we have shown in this report that highly degenerate oligonucleotides can be used for homology prob-

### Table 1. Deduced amino acid and nucleotide sequence identities of the SDH small subunit.

<table>
<thead>
<tr>
<th></th>
<th>Human</th>
<th>Rat</th>
<th>Bovine</th>
<th>D. melanogaster</th>
<th>A. thaliana</th>
<th>S. pombe</th>
<th>S. cerevisiae</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Human</strong></td>
<td>100</td>
<td>99.3</td>
<td>95.3</td>
<td>96.6</td>
<td>63.5</td>
<td>62.9</td>
<td>68.7</td>
</tr>
<tr>
<td><strong>Rat</strong></td>
<td>98.4</td>
<td>100</td>
<td>94.8</td>
<td>95.9</td>
<td>65.6</td>
<td>63.9</td>
<td>70.8</td>
</tr>
<tr>
<td><strong>Bovine</strong></td>
<td>—</td>
<td>92.2</td>
<td>100</td>
<td>66.2</td>
<td>63.9</td>
<td>71.4</td>
<td>—</td>
</tr>
<tr>
<td><strong>D. melanogaster</strong></td>
<td>98.4</td>
<td>97.3</td>
<td>—</td>
<td>100</td>
<td>63.3</td>
<td>62.9</td>
<td>68.9</td>
</tr>
<tr>
<td><strong>A. thaliana</strong></td>
<td>61.9</td>
<td>63.0</td>
<td>—</td>
<td>62.6</td>
<td>100</td>
<td>98.6</td>
<td>75</td>
</tr>
<tr>
<td><strong>S. pombe</strong></td>
<td>61.5</td>
<td>62.8</td>
<td>—</td>
<td>61.8</td>
<td>97.5</td>
<td>100</td>
<td>74.8</td>
</tr>
<tr>
<td><strong>S. cerevisiae</strong></td>
<td>65.6</td>
<td>66.3</td>
<td>—</td>
<td>65.5</td>
<td>65.5</td>
<td>66.0</td>
<td>100</td>
</tr>
</tbody>
</table>

In the upper right of the table are the percent amino acid identities for each pair of proteins. In the lower left are the percent nucleotide identities for each pair of genes.
An important question regarding this technique is whether it is useful only for the isolation of the SDH IP gene or if it can be applied to the isolation of other genes. Recent results from our and other laboratories have demonstrated that it can be used to isolate homologues of the gene encoding the protein kinase subunit of maturation-promoting-factor (S.J.G., unpublished observations) as well as for the isolation of the gene encoding the cyclic-AMP-regulated transcription factor CREB (16). Thus, this technique should have general applications beyond the particular example presented in this report. Furthermore, the target sequence can be present at relatively low abundance, as seen from the success with the use of genomic DNAs for cloning the SDH IP gene; our preliminary results also show that the cDNA for the human SDH IP is present at a very low abundance in the cDNA library used for the cloning (K. Carine and I.E.S., unpublished observations).

Though it is clear that the polymerase chain reaction offers many potential advantages for the cloning of genes based on amino acid sequence information, we do not wish to leave the impression that there are no drawbacks to the technique. First of all, it is at present possible to synthesize only partial cDNA or genomic clones of the gene of interest. The polymerase chain reaction product must subsequently be used as a probe to identify a full-length clone, if that is the desired goal. Also, this procedure currently requires amino acid sequence information from two regions of a protein, or two partially known regions of homology in the target gene, reducing its usefulness in instances where only one conserved region exists in a protein of interest. However, it may be possible to use primers directed to DNA sequences in library vectors that flank the site at which clones have been inserted, allowing one to amplify clones on the basis of knowledge of only a single stretch of amino acids.

Another point to consider is the advantages and drawbacks of using this technique for cloning genes which are members of larger families and subfamilies. On the other hand, the method may be useful for the identification of members of gene families (S.J.G., unpublished observations). On the other hand, if the cloning of a specific member of a gene family is desired, it may be necessary to characterize many additional members of the gene family before isolating the particular clone of interest. In our case, we believe that the SDH IP gene is not a representative of a large gene family. In fact, there is only a single gene in *S. cerevisiae* (A. Lombardo and I.E.S., unpublished observations).

The application of the polymerase chain reaction described in this report may be particularly useful to those engaged in studies of molecular evolution. In the past the necessary protein and DNA sequence data have often been collected haphazardly, depending on the general interest in a particular protein or organism. With the procedure described here it will be feasible to systematically collect sequence information from a wide variety of organisms, even if these are uncommonly studied and genetic information about them is scarce. In many cases it will be possible to use genomic DNA directly, without the need for the construction of genomic or cDNA libraries. An additional advantage may be gained when amplification by the polymerase chain reaction is coupled directly to the sequencing of the amplified product (17). In this way potential "sampling" artifacts resulting from errors introduced by the Taq DNA polymerase or from the propagation of single clones in *E. coli* may be avoided. In our case, uncertainties due to errors by the Taq DNA polymerase seemed to be minimal. The nucleotide sequences of two human clones gave predicted amino acid sequences that were identical and were also highly homologous (95.3% identity) with the independently derived amino acid sequence of the analogous bovine protein. The same observation applies to the rat SDH IP clones (94.8% identity with bovine SDH IP).
The DNAs cloned so far represent about two-thirds of the coding sequence for the SDH IP subunit. The selection of full-length cDNA clones from human and yeast libraries is the logical next step. However, the partial clone from yeast has already been useful in the construction of SDH-deficient \textit{S. cerevisiae} mutants by the method of targeted gene disruption (18). This result provides additional confirmation that we cloned the gene encoding the SDH IP subunit. An SDH-deficient mammalian cell mutant has been described by our laboratory (19, 20), and it will soon be possible to carry out complementation tests using the full-length human gene. We expect the cloned genes and the available mutants to be very useful tools in the study of the mechanisms and the regulation of the assembly of the multisubunit complex II which includes SDH.

Though it may not be prudent to deduce evolutionary relationships on the basis of data from a single gene, it seems clear that the amino acid sequence similarities among the SDH IP subunits from various species are mostly consistent with currently accepted taxonomic hierarchies. As expected from previous data for the bacterial and bovine proteins (11, 13), the IP subunit of SDH is a highly conserved protein. A sequence of 22 amino acids (residues 151–172) is absolutely identical in all eukaryotes, and differs only in one amino acid in \textit{E. coli}. All but the prokaryotic protein are also identical in the amino acids 178–190. The region encompassing amino acids 151–172 contains multiple cysteine residues which are likely to be participants in the formation of certain iron–sulfur clusters, and their conservation would have been fully expected. It is noteworthy that the \textit{D. melanogaster} gene and protein (Fig. 2 and Table 1) differ by less than 5% from the mammalian sequences. Where the other species differ from the vertebrates and insect, the \textit{S. pombe} and \textit{A. thaliana} amino acid sequences are identical (residues 146–150, 175–177, and 200–203).

The entire partial sequence of the human SDH IP subunit was used to search for related amino acid sequences in the PIR protein sequence databank (version 56.0). Not surprisingly, over a 148 amino acid overlap there was a significant identity (25.7%) with fumarate reductase (EC 1.3.99.1) from prokaryotes, an enzyme that catalyzes the reverse reaction of SDH (converts fumarate to succinate). The search also uncovered homologies to a number of ferredoxins. Typically there is 40–55% identity to the ferredoxins in the sequence C I L C A C C S T S C P (amino acids 158–169) and in all cases there was a 100% match to the sequence C X C X C X X X X X C P. A search of the same protein data bank for proteins containing the conserved motif C X C X C X X X X X C P identified additional ferredoxins not identified in the previous screen, and it also identified a number of other proteins with iron–sulfur clusters such as the frxA and frxB proteins of chloroplasts, an isoform of the B-chain of formate dehydrogenase (EC 1.2.1.2) of \textit{Methanobacterium formicicum}, and the hydrogenase (EC 1.18.99.1) of \textit{Desulfovibrio vulgaris}. These data support the previous suggestion that the iron–sulfur clusters in the SDH IP subunit are derived from the ferredoxins (13). However, outside of these clusters the relationship between these proteins is not apparent, and more sophisticated analyses may have to be performed.

In summary, we have demonstrated several distinct advantages of the DNA polymerase chain reaction for homology probing as compared to filter hybridization. The most notable of these is the ability to use highly degenerate oligonucleotides as probes. We have used this approach to homology probing to identify and isolate cDNA or genomic clones of the gene for the IP subunit of SDH from a wide variety of species. Amino acid sequence comparisons among these isolates of the SDH IP subunit and with other proteins reveal that a peptide with the sequence C X C X C X X X C P may be a common element of proteins that contain iron–sulfur clusters. In addition to its utility in cloning evolutionarily conserved genes, this powerful technique should also be applicable to the identification of novel members of gene families.

We thank Hiroto Okayama, Bill Dunphy, Nicolas Brown, Louise Clarke, Mike Yaffe, Nigel Crawford, and Joe Dynes for their generous gifts of libraries and genomic DNAs. Special thanks are extended to Russell Doolittle for his thoughtful comments on this manuscript. This work was supported by a U.S. Public Health Service grant to I.E.S. (GM 33752). S.J.G. was supported by a fellowship from the Powell Foundation.