Recent Advances in the Polymerase Chain Reaction

HENRY A. ERLICH, DAVID GELFAND, JOHN J. SNINSKY

The polymerase chain reaction (PCR) has dramatically altered how molecular studies are conducted as well as what questions can be asked. In addition to simplifying molecular tasks typically carried out with the use of recombinant DNA technology, PCR has allowed a spectrum of advances ranging from the identification of novel genes and pathogens to the quantitation of characterized nucleotide sequences. PCR can provide insights into the intricacies of single cells as well as the evolution of species. Some recent developments in instrumentation, methodology, and applications of the PCR are presented in this review.

SINCE ITS INTRODUCTION IN 1985, THE POLYMERASE CHAIN reaction (PCR) (*1, 2*) has transformed the way DNA analysis is carried out in both research and clinical laboratories. The PCR, which was developed by scientists at Cetus, involves the in vitro enzymatic synthesis of millions of copies of a specific DNA segment. The reaction is based on the annealing and extension of two oligonucleotide primers that flank the target region in duplex DNA; after denaturation of the DNA, each primer hybridizes to one of the two separated strands such that extension from each 3′ hydroxyl end is directed toward the other. The annealed primers are then extended on the template strand with a DNA polymerase. These three steps (denaturation, primer binding, and DNA synthesis) represent a single PCR cycle. Although each step can be carried out at a discrete temperature (for example, 94° to 98°C, 37° to 65°C, and 72°C, respectively), a reaction cycled between the denaturation and the primer binding temperatures generally allows sufficient time for polymerase activity to amplify short PCR products. If the newly synthesized strand extends to or beyond the region complementary to the other primer, it can serve as a primer binding site and template for a subsequent primer extension reactions. Consequently, repeated cycles of denaturation, primer annealing, and primer extension result in the exponential accumulation of a discrete fragment whose termini are defined by the 5′ ends of the primers. This exponential amplification results because under appropriate conditions the primer extension products synthesized in cycle "n" function as templates for the other primer in cycle "n + 1." The length of the products generated during the PCR is equal to the sum of the lengths of the two primers plus the distance in the target DNA between the primers. PCR can amplify double- (ds) or single-stranded (ss) DNA, and with the reverse transcription of RNA into a cDNA copy, RNA can also serve as a target. Because the primers become incorporated into the PCR product and mismatches between the primer and the original genomic template can be tolerated, new sequence information (specific mutations, restriction sites, regulatory elements) and labels can be introduced via the primers into the amplified DNA fragment (*3–5*).

The ability to amplify as well as modify a specific target DNA sequence from a complex template in a simple automated procedure has facilitated many tasks in molecular biology research (for example, cloning and sequencing), thus opening up new areas for experimental investigation. Here, we review some of the recent developments in PCR procedures as well as some recent applications.

New Procedures and Reagents

The initial studies that developed the PCR (*1, 2*) utilized the Klenow fragment of *Escherichia coli* DNA polymerase I to amplify specific targets from human genomic DNA. The inactivation of this polymerase at the high temperatures necessary for strand separation required the addition of enzyme after the denaturation step of each cycle. This rather tedious step was eliminated by the introduction of a thermostable DNA polymerase, the Taq DNA polymerase, isolated from the thermophilic bacterium *Thermus aquaticus* (*6*). The use of Taq DNA polymerase has transformed the PCR by allowing the development of simple automated thermal cycling devices for carrying out the amplification reaction in a single tube containing the necessary reagents (7). The availability of a thermostable enzyme has not only simplified the procedure for the PCR but has increased the specificity and yield of the amplification reaction (7). The incorporation of Taq DNA polymerase into the PCR protocol allows the primers to be annealed and extended at much higher temperature than was possible with Klenow fragment, eliminating much of the nonspecific amplification. Moreover, long PCR products could be amplified from genomic DNA, probably due to a reduction in the secondary structure of the template strands at the elevated temperature used for primer extension. The upper size limit for amplification with the Klenow fragment was only about 400 bp. Fragments as large as 10 kb have been synthesized with Taq DNA polymerase and other thermostable enzymes.

Recent developments in PCR amplification protocols have affected critical parameters, including the misincorporation rate, specificity (target versus nontarget amplification), and maximum length of PCR products. Taq DNA polymerase has no 3′ to 5′ exonuclease ("proofreading") activity, but has a 5′ to 3′ exonuclease activity during polymerization. The initial estimates of the misincorporation rate by Taq DNA polymerase during PCR (about 10^{-4} nucleotides per cycle) were based on measuring the frequency of nucleotide substitutions in the sequence analysis of cloned PCR products (7). Since then, changes in the PCR conditions such as lower concen-

The authors are in the Department of Human Genetics, Core Technology, and Department of Infectious Diseases, Cetus Corporation, Emeryville, CA 94608.

trations of dNTP's and $MgCl_2$, higher annealing temperatures, and shorter extension times have reduced the misincorporation rate to less than 10^{-5} nucleotides per cycle (*8, 9*).

Other DNA polymerases, like the bacteriophage T4 DNA polymerase, appear to have a very low misincorporation rate in PCR (*10*). The heat lability of this enzyme, however, limits its utility. The thermostable VENT DNA polymerase, an enzyme recently isolated from *Thermococcus litoralis*, has the 3′ to 5′ exonuclease activity, and may therefore have a lower misincorporation rate. Further studies will be required to determine if the capacity of this 3′ to 5′ exonuclease activity to degrade single-stranded molecules (like oligonucleotide primers or PCR product prior to primer annealing) will pose problems for PCR amplification. In addition, DNA polymerases with 3′ to 5′ exonuclease activity probably cannot be used in sequence-specific priming reactions because this activity removes the mismatched base at the 3′ end of the primer.

For most PCR applications, such as direct sequencing (*11, 12*) or oligonucleotide probe typing (*13*), it is the population of amplification products that is analyzed, and therefore rare misincorporated nucleotides are not detected. A specific misincorporation will be present in a significant fraction of the PCR products only if the insertion of an incorrect base occurs at the first cycle of a reaction initiated with very few templates. The determination of individual cloned sequences derived from PCR products, however, can reveal such rare errors; the analysis of the PCR products by denaturing gradient gel electrophoresis has also been used to reveal the presence of sequences that contain in vitro mutations (*10*).

Another category of rare PCR artifacts revealed initially by sequencing cloned PCR amplification products is the hybrid sequences generated by in vitro recombination or template strand switching (the so-called "shuffle clones") (*7, 14, 15*). Although misincorporation is rare and the appearance of hybrid sequences even less frequent, it is advisable to sequence multiple clones derived from a single amplification reaction or to clone the PCR products from several independent amplifications to distinguish the sequence of the true genomic template from any potential PCR artifact. In amplification reactions from heterozygous individuals, hybrid sequences can result from a primer that has been partially extended on one allelic template switching to the other allelic template in a subsequent cycle. In the amplification of multigene families, the potential for switching to another template is increased. The probability of partially extended primers competing successfully with the primer is a function of the PCR product concentration. Two recent studies on this PCR artifactual recombination may therefore have exaggerated its significance for genomic DNA amplification by studying systems with very high concentrations of plasmid target DNA sequences (*14, 15*).

The probability that a primer will not be fully extended (will not reach the other primer binding site) is dependent on distance, secondary structure, extent of DNA degradation in the template (*14*), enzyme limitation, the time allowed for polymerase extension, and the processivity of the polymerase. The use of DNA polymerases with increased processivity as well as conditions that increase processivity (for example, accessory proteins or lower salt concentrations) may aid in achieving full primer extension. More importantly, new DNA polymerases may allow the amplification of larger PCR products, a development that would be valuable for the physical mapping and sequencing aspects of the Human Genome Project (*16*). Richardson and co-workers (*17*) have shown that a 12-kD *Escherichia coli* protein can increase the affinity of the bacteriophage T7 DNA polymerase for a primer-template complex, conferring a higher degree of processivity on the enzyme. However, given the thermal cycling properties of the PCR, only enzymes and auxiliary proteins that are heat-resistant are likely to function in new PCR strategies.

Genetically engineered variants of the thermostable Taq DNA polymerase exhibit properties that can be useful for amplifying longer PCR products. For example, a mutant Taq DNA polymerase lacking the 5′ to 3′ exonuclease (*18*) permits efficient amplification of long fragments. A truncated Taq DNA polymerase (the Stoffel fragment) that lacks the 5′ to 3′ exonuclease has been generated and may reveal properties valuable for particular PCR applications.

In general, enzymes that have or lack the 5′ to 3′ and 3′ to 5′ exonuclease activities will perform differently in the PCR. The 5′ to 3′ exonuclease of Taq DNA polymerase has been exploited in a detection assay based on the cleavage of an oligonucleotide probe labeled at the 5′ end and blocked at the 3′ end. The labeled oligonucleotide hybridizes to a target sequence 3′ to one primer. In the presence of amplified target DNA, the exonuclease activity of the polymerase cleaves a labeled fragment from the 5′ end of the annealed oligonucleotide during primer elongation (*19*). What results is the release of a sequence-specific signal concomitant with amplification.

Many of the new thermostable polymerases have additional useful activities. The thermostable DNA polymerase from *Thermus thermophilus* (*Tth*) can reverse-transcribe RNA efficiently in the presence of $MnCl_2$ at high temperatures (*20*). The DNA polymerase activity is enhanced by chelating Mn^{2+} and adding $MgCl_2$, allowing the cDNA synthesis and PCR amplification to be carried out in a single-enzyme, single-tube reaction. In addition, these new thermostable polymerases may be more resistant to blood components that inhibit Taq DNA polymerase.

Reaction Specificity

The initial PCR amplifications with the Klenow fragment were not highly specific; although a unique DNA fragment could be amplified ~200,000-fold from genomic DNA, only about 1% of the PCR product was the targeted sequence (*3*). A specific hybridization probe was therefore required to analyze the amplified DNA (*1, 13*). Amplification with the Taq DNA polymerase greatly increased the specificity of the reaction (*7*), so that for many amplifications the PCR products could be detected as a single ethidium bromide–stained band on an electrophoretic gel. Conditions that increase the stringency of primer hybridization, such as higher annealing temperatures and lower $MgCl_2$ concentrations, enhance specific amplification. In addition, the concentration of enzyme and primers as well as the annealing time, extension time, and number of cycles also effect the specificity of the PCR. The concentration of a specific sequence in a sample can also influence the relative homogeneity of the PCR products. Thus, a single copy nuclear gene present twice in every diploid cell can usually be detected as a unique band after gel electrophoresis of the PCR products; amplification of a sequence present in only one of 10,000 cells is likely to yield a more heterogenous gel profile.

New approaches to improve specificity have been developed. These strategies are based on the recognition that the Taq DNA polymerase retains considerable enzymatic activity at temperatures well below the optimum for DNA synthesis. Thus, in the initial heating step of the reaction, primers that anneal nonspecifically to a partially single-stranded template region can be extended and stabilized before the reaction reaches 72°C for extension of specifically annealed primers. Some of these nonspecifically annealed and extended primers may be oriented with their 3′ hydroxyl directed toward each other, resulting in the exponential amplification of a nontarget fragment. If the DNA polymerase is activated only after the reaction has reached high (>70°C) temperatures, nontarget

amplification can be minimized (*21*). This can be accomplished by manual addition of an essential reagent (DNA polymerase, magnesium chloride, primers) to the reaction tube at elevated temperatures, an approach termed "hot start." The addition of *E. coli* ssDNA binding protein has also been reported to increase specific amplification (*22*).

Hot start not only improves specificity but minimizes the formation of the so-called "primer-dimer," a double-stranded PCR product consisting of the two primers and their complementary sequences. This designation may be somewhat misleading, as sequence analysis of some of these products indicates that additional bases are inserted between the primers. As a result, a fraction of these artifacts may be due to spurious nonspecific amplification of similar but distinct primer binding regions that are positioned in the immediate vicinity. The synthesis of these nonspecific products appears to be a function of the primer concentration. The initial formation of these products can occur at low (ambient) temperatures; hot start eliminates this phase of the thermal profile. Single internal nucleotide substitutions in one of the primers may lead to differences in the size and amount of these background products. Once formed, this artifactual PCR product is efficiently amplified; it is often detected in reactions with rare or no specific template. In fact, the presence of this artifact can be used to distinguish reactions that lack template from unsuccessful amplifications that result from an inhibitor of Taq DNA polymerase. In rare template reactions, these short amplification products may compete with the target fragment for primers and enzyme and prevent efficient target amplification.

Another approach that can improve PCR specificity is to follow the initial amplification reaction with an additional PCR with internal, single, or double nested primers. Like the use of oligonucleotide hybridization probes, this approach utilizes sequence information internal to the two outer primers to identify the subset of amplification products that corresponds to the target fragment. The first use of a nested primer strategy was in the amplification of β-globin from human DNA with the Klenow fragment (*2*). One complication of the conventional nesting strategy is that it requires opening the reaction tube, eliminating or decreasing the concentration of the original outer primers, and adding the inner primers. Although this can sometimes be accomplished by simply diluting the first reaction into a second one, this procedure is both inconvenient and risky, in that it provides an opportunity for contaminating the secondary reaction with amplified product from previous reactions. This problem can be overcome if the outer and inner primers are all present in the initial reaction mix and if the thermal profile is programmed to allow the outer primers but not the inner primers to amplify initially and then to allow the inner but not the outer primers to amplify the targeted subset of the initial PCR products. This can be accomplished with an outer primer that has a "GC-clamp" at the 5′ end such that the melting temperature of the initial PCR product is high, and with an inner primer that is sufficiently short (or AT-rich) so that it cannot bind to the template at the annealing temperature used in the initial PCR cycling. If the initial PCR conditions involve a thermal profile with elevated denaturation and annealing temperatures, then the outer pair of primers but not the inner primers will function. After n cycles, the annealing temperature is dropped, allowing the inner primer or primers to anneal. Then, after $n + x$ cycles the denaturation temperature is dropped, preventing the initial PCR product from denaturing and serving as a template and a binding site for the outer primers. This strategy, termed "drop-in, drop-out nested priming," has been carried out with a single inner nested primer (hemi-nesting) (*23*), resulting in the amplification of a single human immunodeficiency virus (HIV) template in the background of genomic DNA from ~70,000 cells (Fig. 1). The "drop-out" of the outer primer can also be accomplished when limiting concentrations of this primer are used, as in the strategy of "asymmetric PCR," a protocol for generating single strands of DNA (*24*).

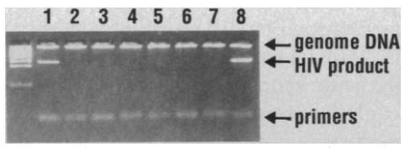

Fig. 1. The use of "drop-in, drop-out" nested primers and "hot start" to enhance specific amplification of single-template HIV sequences. Single-template PCR reactions were set up by limiting dilution of cloned HIV *gag* sequences into human genomic DNA (0.5 μg). The expectation of the Poisson distribution for the concentration used (~0.5 templates per tube) is that four of eight tubes would have one template. The use of a thermal profile that dropped the annealing temperature by 12°C after 29 (n) cycles and decreased the denaturation temperature by 10°C after 31 ($n + 2$) cycles allowed single-tube nested primer amplification. Primer extension was initiated by adding the primers after the temperature of the reaction had reached 72°C. Details of the reaction and the sequences of the three primers were as described (*23*). Lanes 1 and 8 show a single ethidium bromide–stained band generated from a single HIV copy present in a highly complex template sample; no other PCR products are observed. No products are observed in the other six reactions, probably as a result of the absence of specific template.

Contamination of PCR Reactions

Because the PCR can generate trillions of DNA copies from a template sequence, contamination of the amplification reaction with products of a previous PCR reaction (product carryover), exogenous DNA, or other cellular material can create problems both in research and diagnostic applications. In general, attention to careful laboratory procedures—pre-aliquoting reagents, the use of dedicated pipettes, positive-displacement pipettes, or tips with barriers preventing contamination of the pipette barrel, and the physical separation of the reaction preparation from the area of reaction product analysis—minimizes the risk of contamination (*25*). Multiple negative controls (no template DNA added to the reaction) are necessary to monitor and reveal contamination. In genetic typing the contamination of a sample reaction can often be detected by a genotyping result with more than two alleles.

Several approaches to minimize the potential for PCR product carryover have been developed, all based on interfering with the ability of the amplification products to serve as templates. One such strategy developed independently by two groups (*26, 27*) utilizes the principles of the restriction-modification and excision repair systems of bacteria to pretreat PCR reactions and selectively destroy DNA synthesized in a previous PCR. In order to distinguish PCR products from sample template DNA, deoxyuridine triphosphate (dUTP) is substituted for deoxythymidine triphosphate (dTTP) in the PCR and is incorporated into the amplification products. The presence of this unconventional nucleotide allows the distinction of products of previous PCR amplifications from the native DNA of the sample. The enzyme uracil *N*-glycosylase (UNG), present in the reaction pre-mix, catalyzes the excision of uracil from any potential single- or double-stranded PCR carry-over DNA present in the reaction prior to the first PCR cycle. RNA can still serve as a template for the PCR because UNG does not excise uracil from RNA. The abasic polynucleotides that result from cleavage by UNG are susceptible to hydrolysis in alkaline solutions (like PCR buffers) and at elevated temperatures. During PCR, the abasic polynucleotides cannot function as templates because of DNA polymerase stalling or strand scission; the aglycosidic linkage is cleaved at the high denaturation temperature of the first PCR cycle. Furthermore, the resulting modified 3′ termini of the degraded carry-over products are incapable of priming if a DNA polymerase that lacks a 3′

exonuclease (such as Taq DNA polymerase) is used. Fortunately, UNG is inactivated at temperatures used in PCR, so that the amplification products generated during the thermal cycling are stable and can accumulate. This elegant strategy for eliminating PCR product carry-over still allows amplified DNA to serve as a target for probe hybridization and for cloning and sequencing.

Another approach to PCR product inactivation involves short-wavelength ultraviolet irradiation of the reaction mixture prior to amplification (*28*). This strategy, however, requires the addition of DNA polymerase and target DNA to the reaction tube after the irradiation-inactivation step. An alternative inactivation method involves the photochemical modification of the amplified DNA, thereby blocking the Taq DNA polymerase from further extension after it encounters a modified base in the template strand. Isopsoralen reagents can be present throughout the reaction and phoactivated after amplification; cyclobutane adducts with pyrimidine bases in the PCR product are formed, preventing subsequent template function (*29*).

Instruments for the PCR and Product Analysis

In addition to the advances in PCR reagents and protocols, new instruments for automated thermal cycling and for analyzing the PCR products have been developed. The reaction vessels accommodated by the first generation of thermal cyclers were standard plastic microfuge tubes. Some new thermal cyclers have increased rates of heating and cooling and of heat transfer to modified reaction vessels (*30*). PCR amplification in capillary tubes allows rapid thermal cycling and has reduced a complete cycle of denaturation, primer annealing, and DNA synthesis to 20 s (*31*); the speed of the temperature changes achieved in this system has allowed the precise definition of temperature optima for each individual step in the PCR cycle. For the more convenient but slower commercial thermal cycling devices, the use of two-temperature (annealing-extension and denaturation) thermal profiles can reduce overall cycling time. The new generation of automated thermal cyclers is also faster than its predecessors, utilizes thin-walled plastic tubes, and can accommodate more samples. Some of these have a reduced thermal gradient across the heating block, resulting in more precise thermal profiles. In addition, the requirement for mineral oil to prevent evaporation and increase the rate of thermal equilibration has been eliminated in some new models. Instruments for the electrophoretic analysis of PCR products labeled with fluorescent primers have also been introduced.

Research Strategies and Applications

As the performance of PCR has improved with advances in amplification protocols, the strategies for applying it to a variety of research problems have increased. Many of these strategies have been developed to overcome one of the apparent limitations of PCR, namely the requirement for specific sequence information to design the amplification primers. Although this requirement still allows the characterization of mutations, polymorphisms, and evolutionary changes in the DNA sequence between the primers, it represents a constraint on the use of PCR to analyze unknown DNA sequences. This limitation has been overcome by a variety of specific strategies. The general approach has been to create primer binding sites by adding DNA of a known sequence and was illustrated initially by the amplification of unknown cDNA sequences cloned into λ gt11 with primers for the vector sequences that flank the insertion site (*7*). In general, flanking sequences for priming are added by ligation (*32*) or homopolymer tailing with terminal transferase (*33*). These approaches, termed anchored or single-sided PCR (*33, 34*), have generally been applied to cDNA, bacterial genomes, or cloned sequences, all of which are less complex than human genomic DNA.

One strategy for the analysis of unknown sequences that flank a region of known sequence (inverse or inside-out PCR) involves the digestion of the template with a restriction enzyme that cuts outside the known sequence (*35–37*). The resulting fragment is then circularized by ligation and amplification is carried out with primers that hybridize to the known sequence but whose 3′ hydroxyls point away from each other. After circularization, the unknown sequences that had flanked the known region become joined between the 3′ termini of the primers and can therefore be amplified and analyzed. A recent approach to capturing flanking sequences that is not dependent on template circularization is to use a specific target primer and a panel of random oligonucleotide primers; specific amplification from the target into the flanking region is detected by a primer for the sequences on one side of the target primer (*38*).

Performing the PCR with generic primers such as those complementary to repetitive DNA families represents another approach to amplifying unknown sequences. Amplification with primers of repetitive DNA sequences dispersed in the human genome has proven valuable in the analysis of yeast artificial chromosome (YAC) clones and somatic cell hybrid cell lines (*39*). Alu, Kpn, and other interspersed repetitive sequences (IRS) have been used for the design of primers and are used either singly, in combination, or with vector sequence primers to generate unique patterns of PCR products (fingerprints). These patterns can be used to construct genome maps from overlapping clones or cell hybrids. Amplification carried out with human IRS-PCR primers can be used to amplify, from templates of limited complexity, DNA fragments that function as hybridization probes or sequencing templates. Fluorescent probes generated by Alu-PCR from hybrid cell lines have been used for in situ hybridization or "chromosome painting" in both metaphase and interphase cells (*40*). PCR amplification with these repetitive sequence primers can also provide a source of the sequence-tagged sites (STS) (*41*) proposed for the integration of various physical and genetic mapping strategies. An approach for sequencing long cloned sequences exploits the ability of transposons to integrate at random into large DNA fragments that have been cloned into vectors (P1 bacteriophage, F′ bacterial plasmids, and YAC vectors) propagated in bacterial and yeast cells. After transposon integration, this elegant strategy then uses primers complementary to the transposon sequences to generate amplified DNA fragments from the cloned inserts (*42, 43*). Transposon PCR strategies may help overcome some of the limitations of IRS-PCR that arise from nonrandom distribution of these repetitive sequence regions. The amplification of DNA fragments from genomic DNA has recently been carried out with short (10-nucleotide) primers of random or arbitrary sequence (*44, 45*). In some cases, the pattern of amplification products, dubbed APPCR (arbitrarily primed PCR) or RAPD (random amplified polymorphic DNA), varies between different individuals and may, thus, serve as potential genetic markers in mapping studies.

In general, these approaches have uncoupled the ability to amplify DNA from the selective extraction of a specific fragment from a complex genomic template, which was a hallmark of the initial PCR strategies. Experimental strategies for isolating a specific targeted sequence can then be applied independently of the PCR amplification. In genomic PCR (*32*), primer sites are ligated onto a restriction enzyme digest of genomic DNA and the resulting DNA fragments are amplified; specific DNA fragments that bind a partic-

ular protein can be identified in electrophoretic mobility shift assays, isolated, and reamplified, thus overcoming the usual losses that accompany purification or enrichment procedures.

Introduction of new sequences via PCR primers. New nucleotide sequence information can be introduced into the PCR product by addition to or alteration of PCR primers; these include restriction enzyme sites as well as specific mutations incorporated for the analysis of structure-function relations (*5*). Specific sequences can be added to the amplified DNA by PCR that are complementary to vector sequences, eliminating the need for restriction enzyme digestion and ligation (*46*). Regulatory elements such as promoters and translation initiation sites can be added to allow expression of the PCR products in in vitro systems (*47*). The isolation of proteins expressed in vitro or in vivo from PCR products has been facilitated by the addition of sequences that encode a monoclonal antibody epitope (*48*), thus allowing one-step affinity purification of the protein encoded by the amplified DNA.

Footprinting (analysis of DNA-protein interaction). The in vitro analysis of protein-DNA interactions known as footprinting involves the determination of a pattern of protection from chemical or enzymatic cleavage conferred by a DNA binding protein onto a specific nucleotide sequence. In vivo footprinting identifies the guanine bases in a DNA segment that are protected from methylation by bound protein. Because only methylated guanosines can be cleaved by piperidine, this approach makes possible the determination of protein-DNA interaction sites within a cell. However, it is difficult to detect these interactions. A modification of the in vivo procedure that utilizes PCR amplification with ligated primer sites to increase the amount of each cleaved genomic fragment has made genomic footprinting much more sensitive and, thus, a highly promising method for studying DNA-protein interactions in vivo (*49*). In addition, the insertion of unconventional bases that modify protein recognition may be useful for delineation of specific DNA-protein contact points.

PCR and the Human Genome Project. The PCR promises to be important in both the mapping (physical and genetic) and the sequencing aspects of the Human Genome Project (*16*). Many physical mapping strategies rely on the use of IRS-PCR to obtain fingerprint patterns of PCR-amplified human chromosome fragments in order to create an overlapping linear array. The sequence tagged site (STS) proposal (*41*) also requires the use of PCR to generate the 200- to 500-bp sites unique to a given genomic fragment, thus identifying the common element necessary to integrate various physical and genetic maps. The STS proposal, which envisions landmark sites distributed ~100 kb apart, eliminates the need to store and distribute clones and probes and allows the dissemination of STS markers in a database for the design of PCR primers.

The ability of PCR to generate large amounts of pure template for sequence analysis will also facilitate the sequencing aspects of the Human Genome Project. PCR can be used to generate single-stranded products to serve as sequencing templates by minor modifications of the amplification protocol (*24, 50–52*). Recently, a variety of protocols for chain-termination sequencing with Taq DNA polymerase have been developed (*53*), including a linear amplification strategy that allows the repeated use of a single template by thermal cycling.

The construction of genetic maps also relies on PCR approaches. Many restriction fragment length polymorphism (RFLP) markers for the existing map consist of two alleles (the presence or absence of a restriction enzyme cleavage site). These markers are less useful than multiallele markers for the fine-structure map because they will not allow discrimination of all four parental chromosomes in many pedigrees. Short repetitive sequences [for example, the $(CA)_n$ dinucleotide repeats] are dispersed every 30 to 60 kb throughout the human genome; many of them exhibit variation in the number of the repeat units (*54*). With primers to unique flanking sequences, these simple repeat regions can be used as polymorphic markers for genetic mapping by measuring the frequency of recombination within human families. Conventional genetic mapping considers each child as the combination of two meiotic products (the sperm and egg) and measures the meiotic recombination between linked markers by pedigree analysis. Thus, the genotype of the meiotic products and the fraction that have undergone genetic exchange are deduced from analysis of the diploid cells of the progeny. The ability of PCR to coamplify linked DNA segments from individual sperm (*55*) has made possible the direct measurement of recombination frequency in the meiotic products themselves. This powerful new approach, termed sperm mapping, can be used to construct detailed genetic maps from a single individual, allowing the analysis of genetic differences in recombination rates. Sperm mapping involves the coamplification of *n* polymorphic sequences from the individual gametes of a donor heterozygous at each of the *n* loci. The amplified products are typed to determine which alleles are present, and the proportion of recombinant haplotypes is calculated. In addition, the analysis of linked loci in individual sperm also offers the opportunity for haplotype assignment without the transitional requirement for pedigree analysis.

Molecular evolution. The nucleotide sequence in DNA provides the most informative set of characters for the reconstruction of the evolutionary history of species, as well as organism or virus phylogeny. A variety of computer programs have been developed for generating phylogenetic trees based on the nucleotide sequences of a given gene from different individuals in various contemporary species. Prior to the introduction of PCR, the difficulty in obtaining so much sequence information prevented the approach of molecular systematics from realizing its full potential. Recently, PCR has been used to amplify mitochondrial DNA sequences from different human populations, allowing the construction of phylogenetic trees that confirm the African origins of humans and suggest a date for the migration out of Africa (*56*). In addition, PCR was used to obtain ribosomal DNA sequences from a variety of fungal species to resolve the phylogenies of the fungi (*57*). PCR has also facilitated the phylogenetic analysis of the origins of polymorphism in the major histocompatibility complex (MHC) by amplifying DNA from a wide range of contemporary primate species (*58*). The capacity of PCR to amplify small amounts of degraded DNA has allowed for the first time the sequence analysis of historic or ancient DNA. Samples of DNA from extinct species like the woolly mammoth and the quagga, a zebra-like animal, have been subjected to PCR amplification (*59, 60*). The MHC sequences from a 7500-year-old mummified brain have also been amplified and sequenced (*61*). In general, museum specimens, such as the dried skin of the quagga or kangaroo rat, can provide a rich source of samples for sequencing and phylogenetic analysis. The ease of obtaining sequence information and the ability to amplify ancient DNA has already had an impact on evolutionary studies, turning museums of natural history into hotbeds of PCR amplification and molecular analyses.

Gene expression. Amplification from mRNA rather than DNA templates allows the analysis of gene expression. This approach requires the conversion of the mRNA template to cDNA with reverse transcriptase and oligonucleotide primers. A thermostable DNA polymerase (from *T. thermophilus*) that uses manganese as a metal cation and can synthesize DNA from both RNA and DNA templates has been isolated; thus, the initial reverse transcription step and the subsequent PCR amplification can be carried out by the same enzyme. The use of this thermostable polymerase with efficient reverse transcriptase activity should reduce the amount of secondary

structure in the mRNA template, allowing more efficient synthesis of the cDNA strand (*26*).

Many studies of gene expression require quantitative analysis of the mRNA transcripts in the cell. Quantitation can pose problems, as the amplification reaction proceeds through an exponential phase to a linear or plateau phase when the amount of polymerase is limiting and when the PCR products reach a concentration where their reassociation competes with primer annealing and extension. Several approaches to quantitative PCR have been developed based on coamplification of a standard template (*62–64*); one method uses an internal standard RNA transcript added to the reaction in known amounts to determine a standard curve. This standard template has the same primer binding sites as the target mRNA but can be distinguished by length or by restriction site analysis from the target PCR product (*63*).

One area of gene expression studies that has benefitted from the use of PCR is the analysis of specific T cell receptor (TCR) mRNA transcripts associated with specific immune responses. Some of these studies have revealed that only a few of the many different potential TCR transcripts are present in specific autoimmune T cell clones (*65*). In humans, only a few monoclonal antibodies are available for detecting TCR expression; consequently, PCR has emerged as the primary analytic method for determining the expressed TCR repertoire. In general, the use of PCR allows the creation of highly specific cDNA libraries, moving the study of gene expression from the level of tissue-related libraries to the level of specific cell lineages.

Identification of mutations and new members of multi-gene families. PCR has proved valuable in the identification of mutations and polymorphisms in amplified DNA. Cloning and sequencing as well as direct sequencing of the PCR products represent the most informative, albeit most laborious, approach to characterizing amplified DNA. Electrophoretic methods such as denaturing gradient gels are sensitive and rapid approaches for revealing sequence variation. Sequence-dependent variation in the electrophoretic mobility of single strands has also been used to identify polymorphisms and mutations [single-strand conformational polymorphism or SSCP] (*66*). Another general approach has been to detect the cleavage products that are generated by the chemical or enzymatic treatment of mismatches between a labeled probe and the PCR product target strand (*67*).

The identification of genes that are related to genes of known sequence has been facilitated by PCR approaches. The use of primers to sequences that are conserved among several members of a gene family can lead to the amplification and isolation of new family members (*68*). These strategies may use degenerate pools of oligonucleotide primers, which consist of oligonucleotides that contain multiple bases at a single position. In some cases, use of inosine (*69*) at sites where all four bases are possible has reduced the multiplicity of the primer pool, thereby increasing the concentration of individual oligonucleotides. This approach has increased the efficiency of the PCR amplification and the success rate of finding new genes.

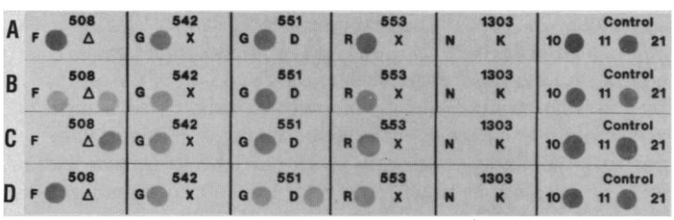

Fig. 2. Reverse dot blot detection of various mutations in the CFTR gene. Biotinylated primer pairs specific for exons 10 and 11 were used to co-amplify a 492-bp and 425-bp fragment, respectively, from DNA samples A to D. The amplified DNA from each sample was hybridized to a CFTR typing strip that contained the array of immobilized probes specific for the normal and the mutant sequence for codons 508, 542, 551, 553, and 1303. (The primers for exon 21 were not used in this experiment, hence the probe dots for the 1303 codon of this exon do not react with the amplified DNA.) Control probes for an invariant sequence in exons 10 and 11 were also immobilized on the nylon membrane strip. The presence of bound PCR product was revealed by incubating the strip with streptavidin-HRP and a chromogenic substrate.

Diagnostic Applications of the PCR

The initial diagnostic application of PCR was in the prenatal diagnosis of sickle cell anemia through the amplification of β-globin sequences. Hybridization of labeled oligonucleotide probes (*1, 13, 70*) or restriction site analysis of the amplified products (*2*) allowed the distinction of normal and mutant alleles. This genetic disease, which is caused by a nucleotide substitution in the β-globin gene, has served as the model system for developing a variety of simple diagnostic methods for detecting a known mutation. The amplification of a specific locus by PCR made the use of nonradioactive allele-specific oligonucleotide (ASO) probes in a dot blot hybridization test a rapid, general, and practical method of genetic typing (*13, 71*). This approach, which depends on the instability of binding of a probe mismatched with the target sequence (*72*), has been applied not only to sickle cell anemia and β-thalassemia (*71*), but to other genetic diseases as well as to the analysis of human leukocyte antigen (HLA) polymorphisms (*73, 74*).

Mismatches between the target sequence and the 3′ base of an annealed oligonucleotide influence the ability of this oligonucleotide to be ligated to another oligonucleotide; this effect forms the basis of the oligomer ligation assay (OLA), a recently automated approach to allelic discrimination (*75*). The specificity of this method has been increased by the introduction of a thermostable DNA ligase isolated from *T. aquaticus* (*76*). This thermostable ligase can be used in conjunction with two pairs of complementary oligonucleotides to generate the target-dependent exponential accumulation of the ligated oligonucleotides (the ligation chain reaction). Another method of allelic discrimination that uses PCR is based on the effect of mismatches in the 3′ end of an oligonucleotide on the priming step (allele-specific amplification). This approach requires that either two separate PCR amplifications be conducted (one with primers specific for the mutation and one with primers specific for the normal allele), or that the PCR products from the two alleles be distinguishable by length or by the label on the two primers (77). This method, which is dependent on the specificity of priming in the first PCR cycle, has recently been applied to sickle cell anemia (*77, 78*).

Unlike sickle cell anemia, most genetic diseases result from a variety of mutations, making the practical diagnostic application of these approaches difficult. For example, the analysis of genetic variation at a locus with n alleles and with n ASO probes would require preparation of n different membranes with the immobilized PCR products from the sample. Each membrane would then be hybridized to a different ASO probe. This problem can be overcome by immobilizing the ASO probes rather than the DNA amplified from the sample. In this "reverse dot blot" method, a membrane that contains an immobilized array of ASO typing probes is hybridized to the PCR product, which is labeled by the incorporation of biotinylated primers (*79*). The presence of the PCR product bound to a specific probe is revealed by incubating the strips with Streptavidin–horseradish peroxidase and a chromogenic substrate. The reverse dot blot approach to genetic disease diagnosis is illustrated by the detection of four of the most frequent cystic fibrosis

mutations (Fig. 2). Cystic fibrosis is the most frequent genetic disease among Caucasians, with a carrier frequency of about 1 in 25. Many different mutations in the cystic fibrosis transport regulator (CFTR) gene have been identified since the isolation of the gene (*80*). Thus, methods like the reverse dot blot, which allow a sample to be analyzed for many different mutations in a single hybridization step, promise to be useful in addressing this and other genetically complex diseases.

When both members of a couple are carriers (heterozygotes containing one mutant allele) for an autosomal recessive disease, they have a one in four chance of having an affected child. For such couples, PCR has made prenatal testing of the fetal amniocytes or the chorionic villus biopsy simpler and faster. The ability to amplify a DNA segment from a single cell has allowed genetic testing to be carried out on in vitro fertilized eggs prior to implantation. A single cell taken from an early developmental stage has been tested; the biopsied embryo was then implanted and developed normally (*81*). In addition, by analyzing one of the haploid cells produced during oogenesis (the polar body), it is sometimes possible to infer the genotype of the egg before it is fertilized (*82*).

In addition to the diagnosis of genetic diseases, PCR has facilitated the analysis of disease susceptibility. Specific alleles at the HLA class I and class II loci are associated with increased risk for a wide variety of autoimmune diseases, such as insulin-dependent diabetes or rheumatoid arthritis (*83*). Virtually all of the extensively polymorphisms at the HLA class II loci (the DRB1 locus has more than 40 alleles) are localized to the second exon, allowing the amplification and analysis of class II sequence variation with primers to conserved regions flanking this exon. This approach, utilizing a panel of oligonucleotide probes in either the dot blot or the reverse dot blot method, forms the basis for a rapid and precise method of HLA typing (*74*). This is valuable not only in the analysis of genetic susceptibility to autoimmune diseases but also in the tissue typing of donor and recipients in transplantation.

In the area of cancer research, PCR has played a role in the identification of chromosomal abnormalities and specific somatic mutations in oncogenes and tumor suppressor genes. Chronic myeloid leukemia (CML) was the first cancer in which a specific genetic abnormality called the Philadelphia chromosome was identified. This chromosomal translocation [t(9;22)] fuses the *abl* and *bcr* genes, resulting in the expression of a chimeric mRNA, the *bcr-abl* transcript. The identification by PCR of the fusion transcripts from this (*84*) and other cancer-associated chromosomal translocations has proved useful in diagnosis as well as in the sensitive detection of minimal residual disease after treatment.

The detection of known mutations in the *ras* oncogenes in a variety of cancers has also been achieved with PCR amplification and dot blot and reverse dot blot hybridization. Some tumors (for example, pancreatic and colon) show a high proportion of *ras* mutations (*85*). Specific mutations in the p53 tumor suppressor gene on chromosome 17 (*86*) as well as the recently described DCC (deleted in colon cancer) locus on chromosome 18 (*87*) and the MCC (missing in colon cancer) locus on chromosome 5 (*88*) have been identified by PCR amplification and sequencing. In addition, p53 mutations were identified in the germline of patients with Li-Fraumeni syndrome (*89*), a hereditary predisposition to breast cancer. For most cancers examined thus far, however, the variety of mutations identified at tumor suppressor loci pose a challenge for simple diagnostic approaches.

Several cancers have been associated with RNA or DNA tumor viruses. The PCR has been used to detect Epstein-Barr virus, human T cell leukemia virus, hepatitis B, and human papilloma virus, which are associated with Burkett's lymphoma and nasopharyngeal carcinoma, adult T cell leukemia, hepatocellular carcinoma, and cervical carcinoma, respectively (*90*). The detection of these viruses, as well as the inherited and somatic mutations and chromosome translocations associated with specific cancers, will provide valuable prognostic and diagnostic information.

The detection of specific pathogen sequences by PCR also promises to be important in infectious disease diagnosis and environmental testing, particularly where the pathogen is difficult to culture. Also, latent viral diseases characterized by few infected cells and transcriptional dormancy will benefit from this approach. The analysis of bacterial, fungal, and parasitic pathogens by PCR will also affect clinical diagnostic practice (*91*). The ability to carry out "retrospective epidemiology" is illustrated in a recent report that identified DNA from *Borrelia burgdorferi* (the Lyme Disease pathogen) in museum specimens of ticks (*59*). In general, the ability to prepare PCR template DNA from a clinical sample (*92*) should have important implications for all diagnostic applications of PCR.

Forensics

In a forensics setting, PCR has allowed the genetic typing of biological evidence found at the crime scene by amplifying polymorphic sequences. DNA fingerprinting, the analysis of fragment length polymorphisms (variable number tandem repeats or VNTR's) based on Southern (DNA) blotting, requires large amounts (>100 ng) of intact, high molecular size DNA. The PCR is capable of analyzing minute quantities of degraded DNA and has been used in the analyses of over 2000 evidence specimens since the first use in 1986 (Pennsylvania versus Pestinikis) (*93*). For most of these evidence samples, DNA typing based on RFLP analysis could not provide a genotyping result. PCR typing of evidence samples for individual identification is based on the analysis of either length polymorphism (VNTR regions) or sequence polymorphisms. The amplification of VNTR regions can be carried out with primers complementary to unique sequences that flank the tandem repeat regions, and allelic variants can be distinguished by determining the length of the PCR products amplified from the sample. However, if the two alleles give rise to PCR products of different lengths, the smaller fragment can also be preferentially amplified. The analysis of degraded DNA can pose problems in that the larger allelic fragment is preferentially lost when amplifying a VNTR polymorphism from a partially degraded DNA sample. Nonetheless, length polymorphisms based on tandem repeats provide a rich source of PCR-based genetic markers for individual identification and genetic mapping (*94, 95*). The amplification of a VNTR in the 16th intron of the retinoblastoma (Rb) gene (*96*) is shown in Fig. 3. The band patterns, revealed by Southern blot analysis, illustrate the preferential amplification of the

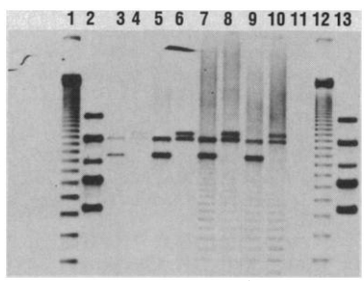

Fig. 3. Amplification of the retinoblastoma VNTR. The amplification of two different human DNA samples was carried out for 24, 26, 28, and 32 cycles with primers and conditions as described (*96*). The agarose gel that contained the PCR products was analyzed by Southern blotting with an HRP-labeled probe complementary to an internal unique sequence (*96*) and the bound probe detected with a chemoluminescent substrate. Lanes 1 and 12, a 123-bp molecular size standard ladder; lanes 2 and 13, a size ladder derived from amplifying β-globin sequences; lanes 3 and 4, the amplification of samples A and B for 24 cycles; lanes 5 and 6, the PCR products after 26 cycles; lanes 7 and 8, after 28 cycles; and lanes 9 and 10, after 32 cycles. Both samples are from heterozygous individuals.

smaller allelic PCR product and the appearance of artifactual "ladder bands" as the cycle number increases. These "ladders," whose "rungs" are separated by the repeat unit length, result presumably from slippage of the polymerase or the out-of-register annealing of repeat unit sequences at high PCR product concentration.

Although sequence polymorphisms can be detected by a variety of methods, in our experience, the use of oligonucleotide hybridization probes represents a general and powerful approach to genetic typing. With the use of the reverse dot blot method, the complexity of the test is independent of the number of probes because the sample is analyzed with all of the probes in a single hybridization reaction. This format has been developed for typing the polymorphism at the HLA-DQA1 locus (*79*) and is commercially available as the first PCR-based genetic typing kit. If the genotype of the evidence sample differs from that of the reference sample of the suspect, the suspect is excluded or eliminated as a potential donor of the evidence. If the genotype of the suspect and evidence sample match, the suspect is included (that is, cannot be eliminated as the donor). The significance of this match for identifying the source of the evidence specimen is a function of the frequency of this genotype as determined by population surveys (*97*). In the >200 cases in which the DQA1 test has yielded conclusive results, 65% have resulted in an inclusion and 35% have resulted in an exclusion (*93*). Similar dot blot tests for determining the sequence polymorphism in the D-loop segment of mitochondrial DNA have been developed and applied to the forensic analysis of bone fragments (*98*).

Summary

The ability to amplify a specific DNA segment in a simple automated reaction has had an enormous impact in many areas. The increased accessibility of DNA analysis made possible by PCR has been referred to as "democratizing the DNA sequence" or enabling the "practice of molecular biology without a permit," in reference to Chargaff's dictum that molecular biology itself was "the practice of biochemistry without a license." As advances in the reagents, instruments, and protocols for carrying out PCR continue, researchers and clinicians will be able to access the information encoded in nucleotide sequences more easily and apply PCR to an even more diverse set of problems.

REFERENCES AND NOTES

1. R. K. Saiki *et al.*, *Science* **230**, 1350 (1985).
2. K. B. Mullis and F. Faloona, *Methods Enzymol.* **155**, 335 (1987).
3. S. J. Scharf, G. T. Horn, H. A. Erlich, *Science* **233**, 1076 (1986).
4. K. B. Mullis *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **51**, 263 (1986).
5. R. Higuchi, in *PCR Technology: Principles and Applications for DNA Amplification*, H. Erlich, Ed. (Stockton, New York, 1989), pp. 61–70.
6. T. D. Brock and H. Freeze, *J. Bacteriol.* **98**, 289 (1969).
7. R. K. Saiki *et al.*, *Science* **239**, 487 (1988).
8. K. A. Eckert and T. A. Kunkel, *Nucleic Acids Res.* **18**, 3739 (1990).
9. D. H. Gelfand and T. J. White, in *PCR Protocols: A Guide to Methods and Applications*, M. A. Innis, D. H. Gelfand, J. J. Sninsky, T. J. White, Eds. (Academic Press, San Diego, CA, 1990), p. 129.
10. P. Keohavang and W. G. Thilly, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9253 (1989).
11. L. A. Wrishnick *et al.*, *Nucleic Acids Res.* **15**, 529 (1987).
12. C. Wong *et al.*, *Nature* **330**, 384 (1987).
13. R. K. Saiki, T. L. Bugawan, G. T. Horn, K. B. Mullis, H. A. Erlich, *ibid.* **324**, 163 (1986).
14. S. Pääbo, D. M. Irwin, A. C. Wilson, *J. Biol. Chem.* **265**, 4718 (1990).
15. A. Meyerhans, J.-P. Vartanian, S. Wain-Hobson, *Nucleic Acids Res.* **18**, 1687 (1990).
16. E. Rose, *FASEB J.* **5**, 46 (1991).
17. H. E. Huber, S. Tabor, C. Richardson, *J. Biol. Chem.* **262**, 16224 (1987).
18. R. D. Abramson, unpublished data.
19. P. M. Holland, R. D. Abramson, R. Watson, D. Gelfand, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
20. T. W. Myers and D. H. Gelfand, *Biochemistry*, in press.
21. F. Faloona, S. Weiss, F. Ferre, K. Mullis, paper presented at the Sixth International Conference on AIDS, 20 to 24 June 1990, San Francisco, CA, 1990.
22. T. R. Holm, unpublished data.
23. R. Higuchi, unpublished data.
24. U. B. Gyllensten and H. A. Erlich, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 7652 (1988).
25. R. Higuchi and S. Kwok, *Nature* **339**, 237 (1989).
26. M. C. Longo, M. S. Berninger, J. L. Harley, *Gene* **93**, 125 (1990).
27. J. J. Sninsky *et al.*, in preparation.
28. G. Sarkar and S. Sommer, *Nature* **343**, 27 (1990).
29. S. T. Isaacs, J. W. Tessman, K. C. Metchette, J. E. Hearst, G. D. Cimino, *Nucleic Acids Res.* **19**, 109 (1991).
30. L. Haff *et al.*, *Biotechniques* **10**, 102 (1991).
31. C. T. Wittwer, G. C. Fillmore, D. J. Garling, *Anal. Biochem.* **186**, 328 (1990).
32. K. W. Kinzler and B. Vogelstein, *Nucleic Acids Res.* **17**, 3645 (1989).
33. E. Y. Loh, J. F. Elliott, S. Cwirla, L. L. Lanier, M. M. Davis, *Science* **243**, 217 (1989).
34. M. A. Frohman, M. K. Duch, G. R. Martin, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8998 (1988).
35. H. Ochman, A. S. Gerber, D. L. Hartl, *Genetics* **120**, 621 (1988).
36. T. Triglia, M. G. Peterson, D. J. Kemp, *Nucleic Acids Res.* **16**, 81 (1988).
37. J. Silver and V. Keerikatte, *J. Virol.* **63**, 1924 (1989).
38. D. Parker *et al.*, *Nucleic Acids Res.*, in press.
39. S. A. Ledbetter, D. L. Nelson, S. T. Warren, D. H. Ledbetter, *Genomics* **6**, 475 (1990).
40. P. Lichter, S. A. Ledbetter, D. H. Ledbetter, D. C. Ward. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 6634 (1990).
41. M. Olson, L. Hood, C. Cantor, D. Botstein, *Science* **245**, 1434 (1989).
42. S. H. Phadnis, H. V. Huang, D. E. Berg, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 5908 (1989).
43. L. D. Strausbaugh, M. T. Bourke, M. T. Sommer, M. E. Coon, C. M. Berg, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
44. J. Welsh and M. McClelland, *Nucleic Acids Res.* **18**, 7213 (1990).
45. J. G. K. Williams, A. R. Kubelik, K. J. Livak, J. A. Rafalski, S. V. Tingey, *ibid.*, p. 6531.
46. C. Aslanidis and P. J. de Jong, *ibid.*, p. 6069.
47. G. Sarkar and S. S. Sommer, *Science,* **244**, 331 (1989).
48. G. A. Martin *et al.*, *Cell* **63**, 843 (1990).
49. P. R. Mueller and B. Wold, *Science* **246**, 780 (1989).
50. R. Higuchi and H. Ochman, *Nucleic Acids Res.* **17**, 5865 (1989).
51. E. Hornes, T. Hultman, T. Moks, M. Uhlen, *Biotechniques* **9**, 730 (1990).
52. G. Ruano and K. K. Kidd, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 2815 (1991).
53. M. D. Brow, in *PCR Protocols: A Guide to Methods and Applications*, M. A. Innis, D. H. Gelfand, J. J. Sninsky, T. J. White, Eds. (Academic Press, San Diego, CA, 1990), p. 189.
54. J. L. Weber and P. E. May, *Am. J. Hum. Genet.* **44**, 388 (1989).
55. H. H. Li *et al.*, *Nature* **335**, 414 (1988).
56. L. Vigilant, R. Pennington, H. Harpending, T. Kocher, A. C. Wilson, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9350 (1989).
57. T. D. Bruns, T. J. White, J. W. Taylor, *Annu. Rev. Ecol. Syst.*, in press.
58. U. B. Gyllensten and H. A. Erlich, *Proc. Natl. Acid. Sci. U.S.A.* **86**, 9986 (1989).
59. D. H. Persing *et al.*, *Science* **249**, 1420 (1990).
60. S. Pääbo, R. Higuchi, A. C. Wilson, *J. Biol. Chem.* **264**, 9709 (1989).
61. D. A. Lawlor, C. D. Dickel, W. Hauswirth, P. Parham, *Nature* **349**, 785 (1991).
62. A. M. Wang, M. V. Doyle, D. F. Mark, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9717 (1989); M. Becker-André and K. Hahlbrock, *Nucleic Acids Res.* **17**, 9347 (1989).
63. G. Gilliland, S. Perrin, K. Blanchard, H. F. Bunn, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 2725 (1990).
64. D. E. Kellogg, J. J. Sninsky, S. Kwok, *Anal. Biochem.* **189**, 202 (1990).
65. H. Acha-Orbea *et al.*, *Cell* **54**, 263 (1988).
66. M. Orita, H. Iwahana, H. Kanazawa, K. Hayashi, T. Sekiya, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2766 (1989).
67. R. G. Cotton, N. R. Rodriguez, R. D. Campbell, *ibid.* **85**, 4397 (1988).
68. D. H. Mack and J. J. Sninsky, *ibid.*, p. 6977; S. I. Girgis *et al.*, *Nucleic Acids Res.* **16**, 10371 (1988); C. C. Lee *et al.*, *Science* **239**, 1288 (1988).
69. The use of modified or unusual bases such as inosine or 5-methylcytosine in oligonucleotide primers and probes provides another variable in strategies to adjust the interaction between oligonucleotides and their target sequences. The enzymatic incorporation of modified bases such as 7-Aza-dGTP (deoxyguanidinetriphosphate) proved useful in amplification through regions with secondary structure resulting from base stacking.
70. S. H. Embury *et al.*, *N. Engl. J. Med.* **316**, 656 (1987).
71. R. K. Saiki *et al.*, *ibid.* **319**, 537 (1988).
72. B. J. Conner *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 278 (1983).
73. H. A. Erlich and T. L. Bugawan, *PCR Technology: Principles and Applications for DNA Amplification*, H. Erlich, Ed. (Stockton, New York, 1989), pp. 193–208.
74. H. Erlich, T. Bugawan, A. Begovich, *Eur. J. Immunogenet.* **18**, 33 (1991).
75. D. A. Nickerson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8923 (1990).
76. F. Barany, *ibid.* **88**, 189 (1991).
77. F. F. Chehab and K. W. Kan, *ibid.* **86**, 9178 (1989).
78. D. Y. Wu, L. Ugozzoli, B. K. Pal, R. B. Wallace, *ibid.* p. 2757.
79. R. Saiki, P. S. Walsh, C. H. Levenson, H. A. Erlich, *ibid.*, p. 6230.
80. J. M. Rommens *et al.*, *Science* **245**, 1059 (1989).
81. A. H. Handyside *et al.*, *Lancet* **i**, 347 (1989).
82. M. Monk and C. Holding, *ibid.* **335**, 985 (1990).
83. G. T. Nepom and H. A. Erlich, *Annu. Rev. Immunol.* **9**, 493 (1991).
84. E. S. Kawasaki *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 5698 (1988).
85. J. L. Bos, *Cancer Res.* **49**, 4682 (1989).
86. J. M. Nigro *et al.*, *Nature* **342**, 705 (1989).

87. E. R. Fearon *et al.*, *Science* **247**, 49 (1990).
88. K. W. Kinzler *et al.*, *ibid.* **251**, 1366 (1991).
89. S. Srivastava, Z. Zou, K. Pirollo, W. Blattner, E. H. Chang, *Nature* **348**, 747 (1990).
90. S. Kwok and J. J. Sninsky, *PCR Technology: Principles and Applications for DNA Amplification*, H. Erlich, Ed. (Stockton, New York, 1989), pp. 235–244.
91. T. J. White, R. Madej, D. H. Persing, *Adv. Clin. Chem.*, in press; S. D. Williams and S. Kwok, in *Laboratory Diagnosis of Viral Infections*, E. H. Lennette, Ed. (Dekker, New York, 1991), pp. 147–173.
92. P. S. Walsh, D. A. Matzger, R. Higuchi, *Biotechniques* **10**, 506 (1991).
93. E. Blake, J. Mihalovich, R. Higuchi, P. S. Walsh, H. Erlich, unpublished data.
94. J. Bill *et al.*, *J. Exp. Med.* **169**, 115 (1989); B. Budowle *et al.*, *Am. J. Hum. Genet.* **48**, 137 (1991).
95. R. C. Allen, G. Graves, B. Budowle, *Biotechniques* **7**, 736 (1989); K. Kasai *et al.*, *J. Forensic Sci.* **35**, 1196 (1990).
96. S. Scharf, unpublished data.
97. R. Helmuth *et al.*, *Am. J. Hum. Genet.* **47**, 515 (1990).
98. M. Stoneking, D. Hedgecock, R. Higuchi, L. Vigilant, H. A. Erlich, *ibid.* **48**, 370 (1991).
99. We thank our colleagues who contributed to the development and application of PCR. The space constraints of this review and the many publications on PCR prevent a comprehensive survey of advances and applications; we apologize to any of our colleagues whose studies have not been noted specifically. We are grateful to R. Saiki, S. Scharf, R. Higuchi, and R. Abramson for allowing us to cite their unpublished work; E. Rose for critical review; and K. Levenson for preparation of this manuscript.

Complementary DNA Sequencing: Expressed Sequence Tags and Human Genome Project

MARK D. ADAMS, JENNY M. KELLEY, JEANNINE D. GOCAYNE, MARK DUBNICK, MIHAEL H. POLYMEROPOULOS, HONG XIAO, CARL R. MERRIL, ANDREW WU, BJORN OLDE, RUBEN F. MORENO, ANTHONY R. KERLAVAGE, W. RICHARD MCCOMBIE, J. CRAIG VENTER*

Automated partial DNA sequencing was conducted on more than 600 randomly selected human brain complementary DNA (cDNA) clones to generate expressed sequence tags (ESTs). ESTs have applications in the discovery of new human genes, mapping of the human genome, and identification of coding regions in genomic sequences. Of the sequences generated, 337 represent new genes, including 48 with significant similarity to genes from other organisms, such as a yeast RNA polymerase II subunit; *Drosophila* kinesin, *Notch*, and *Enhancer of split*; and a murine tyrosine kinase receptor. Forty-six ESTs were mapped to chromosomes after amplification by the polymerase chain reaction. This fast approach to cDNA characterization will facilitate the tagging of most human genes in a few years at a fraction of the cost of complete genomic sequencing, provide new genetic markers, and serve as a resource in diverse biological research fields.

THE HUMAN GENOME IS ESTIMATED TO CONSIST OF 50,000 to 100,000 genes, up to 30,000 of which may be expressed in the brain (*1*). However, GenBank lists the sequence of only a few thousand human genes and <200 human brain messenger RNAs (mRNAs) (*2*). Once dedicated human chromosome sequencing begins in 5 years, it is expected that 12 to 15 years will be required to complete the sequence of the genome (*3*). It is therefore likely that the majority of human genes will remain unknown for at least the next decade. The merits of sequencing cDNA, reverse transcribed from mRNA, as a part of the human genome project have been vigorously debated since the idea of determining the complete nucleotide sequence of humans first surfaced. Proponents of cDNA sequencing have argued that because the coding sequences of genes represent the vast majority of the information content of the genome, but only 3% of the DNA, cDNA sequencing should take precedence over genomic sequencing (*4*). Proponents of genomic sequencing have argued the difficulty of finding every mRNA expressed in all tissues, cell types, and developmental stages and have pointed out that much valuable information from intronic and intergenic regions, including control and regulatory sequences, will be missed by cDNA sequencing (*5*). However, many genome enthusiasts have incorrectly stated that gene coding regions, and therefore mRNA sequences, are readily predictable from genomic sequences and have concluded that there is no need for large-scale cDNA sequencing. In fact, prediction of transcribed regions of human genomic sequence is currently feasible only for relatively large exons (*6*).

On the basis of our high output with automated DNA sequence analysis of 96 templates per day and consideration of the above issues, we initiated a pilot project to test the use of partial cDNA sequences (ESTs) in a comprehensive survey of expressed genes.

Sequence-tagged sites (STSs) are becoming standard markers for the physical mapping of the human genome (*7*). These short sequences from physically mapped clones represent uniquely identified map positions. ESTs can serve the same purpose as the random genomic DNA STSs and provide the additional feature of pointing directly to an expressed gene. An EST is simply a segment of a sequence from a cDNA clone that corresponds to an mRNA. ESTs longer than 150 bp were found to be the most useful for similarity searches and mapping.

M. D. Adams, J. M. Kelley, J. D. Gocayne, M. Dubnick, A. Wu, B. Olde, R. F. Moreno, A. R. Kerlavage, W. R. McCombie, and J. C. Venter are in the Section of Receptor Biochemistry and Molecular Biology, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892. M. H. Polymeropoulos, H. Xiao, and C. R. Merril are in the Laboratory of Biochemical Genetics, National Institute of Mental Health, Neuroscience Center at St. Elizabeth's Hospital, Washington, DC 20032.

*To whom correspondence should be addressed.