The Unusual Origin of the Polymerase Chain Reaction

A surprisingly simple method for making unlimited copies of DNA fragments was conceived under unlikely circumstances—during a moonlit drive through the mountains of California

by Kary B. Mullis

Sometimes a good idea comes to you when you are not looking for it. Through an improbable combination of coincidences, naiveté and lucky mistakes, such a revelation came to me one Friday night in April, 1983, as I gripped the steering wheel of my car and snaked along a moonlit mountain road into northern California's redwood country. That was how I stumbled across a process that could make unlimited numbers of copies of genes, a process now known as the polymerase chain reaction (PCR).

Beginning with a single molecule of the genetic material DNA, the PCR can generate 100 billion similar molecules in an afternoon. The reaction is easy to execute: it requires no more than a test tube, a few simple reagents and a source of heat. The DNA sample that one wishes to copy can be pure, or it can be a minute part of an extremely complex mixture of biological materials. The DNA may come from a hospital tissue specimen, from a single human hair, from a drop of dried blood at the scene of a crime, from the tissues of a mummified brain or from a 40,000-year-old woolly mammoth frozen in a glacier.

In the seven years since that night, applications for the PCR have spread throughout the biological sciences: more than 1,000 reports of its use have been published. Given the impact of the PCR on biological research and its conceptual simplicity, the fact that it lay unrecognized for more than 15 years after all the elements for its implementation were available strikes many observers as uncanny.

The polymerase chain reaction makes life much easier for molecular biologists: it gives them as much of a particular DNA as they want. Casual discussions of DNA molecules sometimes make them sound like easily obtained objects. The truth is that in practice it is difficult to get a well-defined molecule of natural DNA from any organism except extremely simple viruses.

The difficulty resides in the nature of the molecule. DNA is a delicate chain made of four deoxynucleotides: deoxyadenylate (*A*), deoxythymidylate (*T*), deoxyguanylate (*G*) and deoxycytidylate (*C*); the sequence of these bases encodes the genetic information. Rarely does one find a single strand of DNA; usually pairs of strands with complementary sequences form double helixes in which the *A*'s in one strand bind with the *T*'s in the other, and the *G*'s bind with the *C*'s [*see illustration on opposite page*]. Inside a cell this DNA helix is surrounded and further coiled by various proteins. When biologists try to isolate a naked DNA chain, the DNA is so long and thin that even mild shearing forces break it at random points along its length. Consequently, if the DNA is removed from 1,000 identical cells, there will be 1,000 copies of any given gene, but each copy will be on a DNA fragment of differing length.

For years this problem made it difficult to study genes. Then in the 1970's enzymes known as restriction endonucleases were discovered: these enzymes snipped strands of DNA at specific points. The endonucleases made it possible to cut DNA into smaller, sturdier, more identifiable pieces and thereby made it easier to isolate the pieces containing a gene of interest.

By the late 1970's, therefore, molecular biologists were busily studying DNA with endonucleases and with other molecules called oligonucleotide probes. An oligonucleotide is a short chain of specifically ordered nucleotide bases. Under the right conditions, an oligonucleotide will bind specifically with a complementary sequence of nucleotides in single-strand DNA. Therefore, radioactively labeled, man-made oligonucleotides can serve as probes for determining whether a sample of DNA contains a specific nucleotide sequence or gene. In 1979 the Cetus Corporation in Emeryville, Calif., hired me to synthesize oligonucleotide probes.

By 1983 the charm of synthesizing oligonucleotides for a living had entered a decline—a decline that most of us so employed were happy to witness. The laborious but very quaint chemical art form for making oligonucleotides manually, to which we had grown comfortably numb, had given way to a much less charming but reliable automated technique. It was an immense improvement.

In the aftermath of this minor industrial revolution, we nucleotide chemists found ourselves success-

KARY B. MULLIS describes himself as "a generalist with a chemical prejudice." In addition to the polymerase chain reaction, he is also known for having invented a plastic that changes color rapidly when exposed to ultraviolet light. While working as a biochemistry graduate student at the University of California, Berkeley, he published a paper in *Nature* entitled "The Cosmological Significance of Time Reversal." Mullis received his Ph.D. in biochemistry in 1972. After working as a postdoctoral fellow at the University of Kansas Medical School and the University of California, San Francisco, Mullis joined the Cetus Corporation, where he discovered the polymerase chain reaction. In 1986 he became the director of molecular biology at Xytronyx, Inc. Today Mullis works in La Jolla, Calif., as a private consultant on polymerase-chain-reaction technology and nucleic acid chemistry.

fully underemployed. Laboratory machines, which we loaded and watched, were making almost more oligonucleotides than we had room for in the freezer and certainly more than the molecular biologists—who seemed to be working even more slowly and tediously than we had previously suspected—could use in their experiments. Consequently, in my laboratory at Cetus, there was a fair amount of time available to think and to putter.

I found myself puttering around with oligonucleotides.

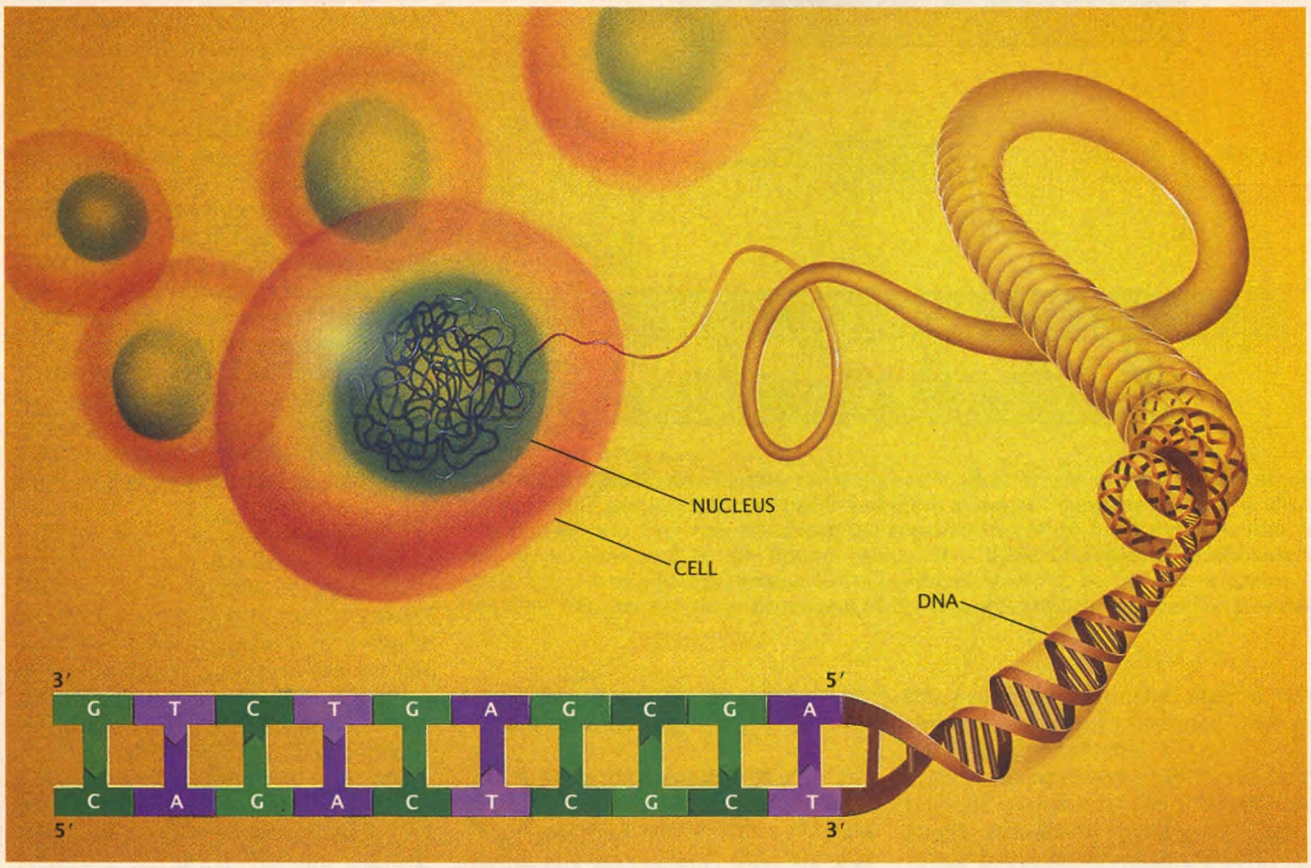
I knew that a technique for easily determining the identity of the nucleotide at a given position in a DNA molecule would be useful, especially if it would work when the complexity of the DNA was high (as it is in human DNA) and when the available quantity of the DNA was small. I did not see why one could not use the enzyme DNA polymerase and a variation of a technique called dideoxy sequencing, and so I designed a simple-minded experiment to test the idea.

To understand the approach I had in mind, it is worth reviewing certain facts about DNA. A strand of the molecule has one end that is known, by chemical convention, as three-prime and one end that is five-prime. In a double helix of DNA, the complementary strands are said to be antiparallel, because the three-prime end of one strand pairs with the five-prime of the other strand, and vice versa.

In 1955 Arthur Kornberg of Stanford University and his associates discovered a cellular enzyme called a DNA polymerase. DNA polymerases serve several natural functions, including the repair and replication of DNA. These enzymes can lengthen a short oligonucleotide "primer" by attaching an additional nucleotide to its three-prime end, but only if the primer is hybridized, or bound, to a complementary strand called the template. The surrounding solution must also contain nucleotide triphosphate molecules as building blocks.

The nucleotide that the polymerase attaches will be complementary to the base in the corresponding position on the template strand. For example, if the adjacent template nucleotide is an *A*, the polymerase attaches a *T* base; if the template nucleotide is a *G*, the enzyme attaches a *C*. By repeating this process, the polymerase can extend the primer's three-prime end all the way to the template's five-prime terminus [*see illustration on page 59*]. In a double helix of DNA, each strand serves as a template for the other during replication and repair.

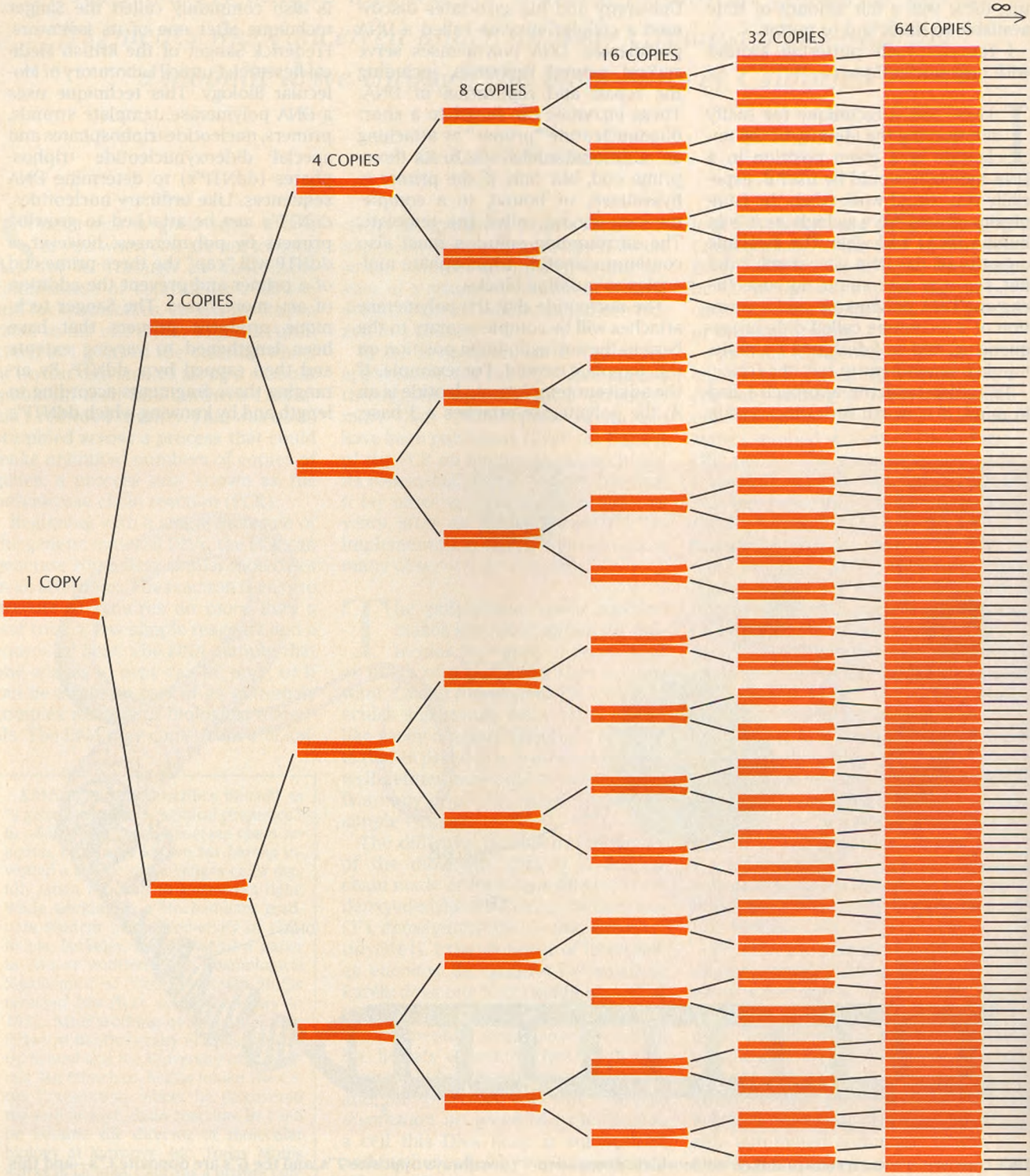
Now for dideoxy sequencing, which is also commonly called the Sanger technique after one of its inventors, Frederick Sanger of the British Medical Research Council Laboratory of Molecular Biology. This technique uses a DNA polymerase, template strands, primers, nucleotide triphosphates and special dideoxynucleotide triphosphates (ddNTP's) to determine DNA sequences. Like ordinary nucleotides, ddNTP's can be attached to growing primers by polymerases; however, a ddNTP will "cap" the three-prime end of a primer and prevent the addition of any more bases. The Sanger technique produces primers that have been lengthened to varying extents and then capped by a ddNTP. By arranging these fragments according to length and by knowing which ddNTP's



DNA consists of two strands of linked nucleotides: deoxyadenylates (*A*'s), deoxythymidylates (*T*'s), deoxyguanylates (*G*'s) and deoxycytidylates (*C*'s). The sequence of nucleotides in one strand is complementary to that in the other strand—the *A*'s are always opposite *T*'s, and the *G*'s are opposite *C*'s—and this complementarity binds the strands together. Each strand has a three-prime and a five-prime end. Because their orientations oppose one another, the strands are said to be antiparallel.

have been added, an investigator can determine the sequence of bases in the template strand. For example, if a dideoxyadenine (ddA) base were added at a given position, the corresponding complementary base in the template would be a *T;* the addition of a dideoxyguanine (ddG) implies the presence of a *C* in the template.

In the modified version of this technique that I was contemplating, I would use only polymerases, templates, ddNTP's and primer molecules—that is, I would omit the ordinary nucleotide triphosphates from the mixture. Extension of the primers would therefore terminate immediately after the addition of one base from a ddNTP to the chain. If I knew which ddNTP had been added to the primers, I would also know the identity of the corresponding base in the template strand. In this way, I could deduce the identity of a base in the template



POLYMERASE CHAIN REACTION is a simple technique for copying a piece of DNA in the laboratory with readily available reagents. Because the number of copies increases exponentially, more than 100 billion can be made in only a few hours.

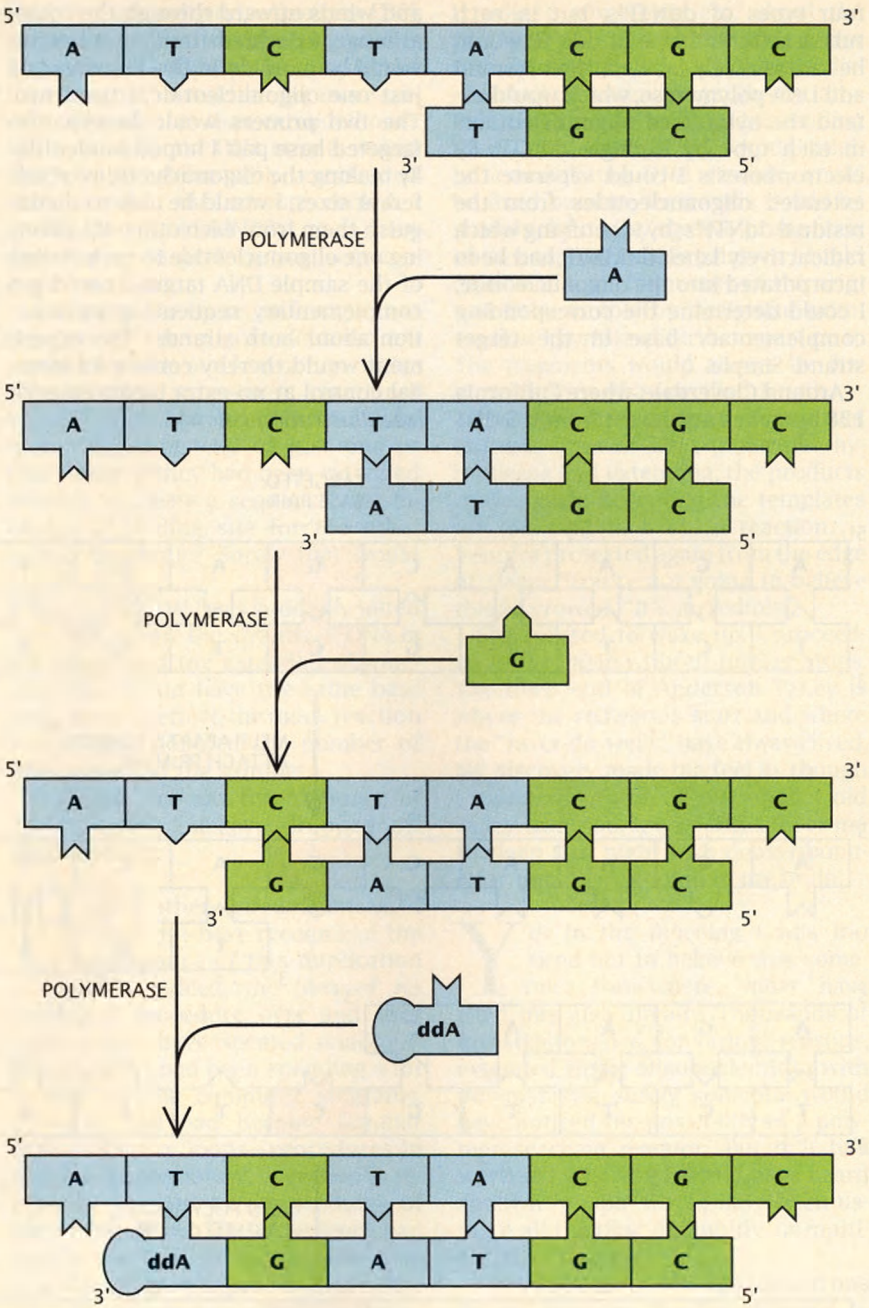
strand adjacent to the site where the primer binds.

What I did not realize at the time was that there were many good reasons why my sequencing idea could not work. The problem was that oligonucleotides sometimes hybridize with DNA sequences other than those intended; these unavoidable pairings would have made my results ambiguous. Even in the hands of those skilled in the art of careful hybridization, it was impossible to bind oligonucleotides to whole human DNA with sufficient specificity to get anything even approaching a meaningful result.

It was because of this limitation that researchers had resorted to more difficult procedures for looking at human DNA. For instance, restriction enzymes could be employed to cleave the DNA sample into various fragments that could be separated from each other by electrophoresis; in this way, the sample could be "purified," to some extent, of all DNA except the target fragment before the hybridization of oligonucleotide probes. This approach reduced erroneous hybridizations sufficiently to provide meaningful data, but just barely. Moreover, this procedure was lengthy and would not work on degraded or denatured samples of DNA.

Another technique that was much too lengthy for routine DNA analysis involved cloning. A human DNA sequence of interest could be cloned, or copied, into a small ring of DNA called a plasmid. Copies of this plasmid and the targeted sequence could then be produced in bacteria, and sequence information could be obtained by oligonucleotide hybridization and dideoxy sequencing. In the early 1980's dideoxy sequencing of cloned DNA was the method by which most human DNA sequence information had been obtained.

In proposing my simple-minded experiment, I was implicitly assuming that no such cloning or other step would be necessary to detect specific human DNA sequences by a single oligonucleotide hybridization. In token defense of my misguided puttering, I can point out that a group down the hall led by Henry A. Erlich, one of Cetus's senior scientists, was trying another method based on the hybridization of a single oligonucleotide to a human DNA target. No one laughed out loud at Henry, and we were all being paid regularly. In fact, we were being paid enough to lead some of us to assume, perhaps brashly, that we were somewhere near the cutting edge of DNA technology.



DNA POLYMERASE, an enzyme, can lengthen a short strand of DNA, called an oligonucleotide primer, if the strand is bound to a longer "template" strand of DNA. The polymerase does this by adding the appropriate complementary nucleotide to the three-prime end of the bound primer. If a dideoxynucleotide triphosphate (ddNTP) such as dideoxyadenine (ddA) is added, however, no further extension is possible, because the three-prime end of the ddA will not link to other nucleotides.

One Friday evening late in the spring I was driving to Mendocino County with a chemist friend. She was asleep. U.S. 101 was undemanding. I liked night driving; every weekend I went north to my cabin and sat still for three hours in the car, my hands occupied, my mind free. On that particular night I was thinking about my proposed DNA-sequencing experiment.

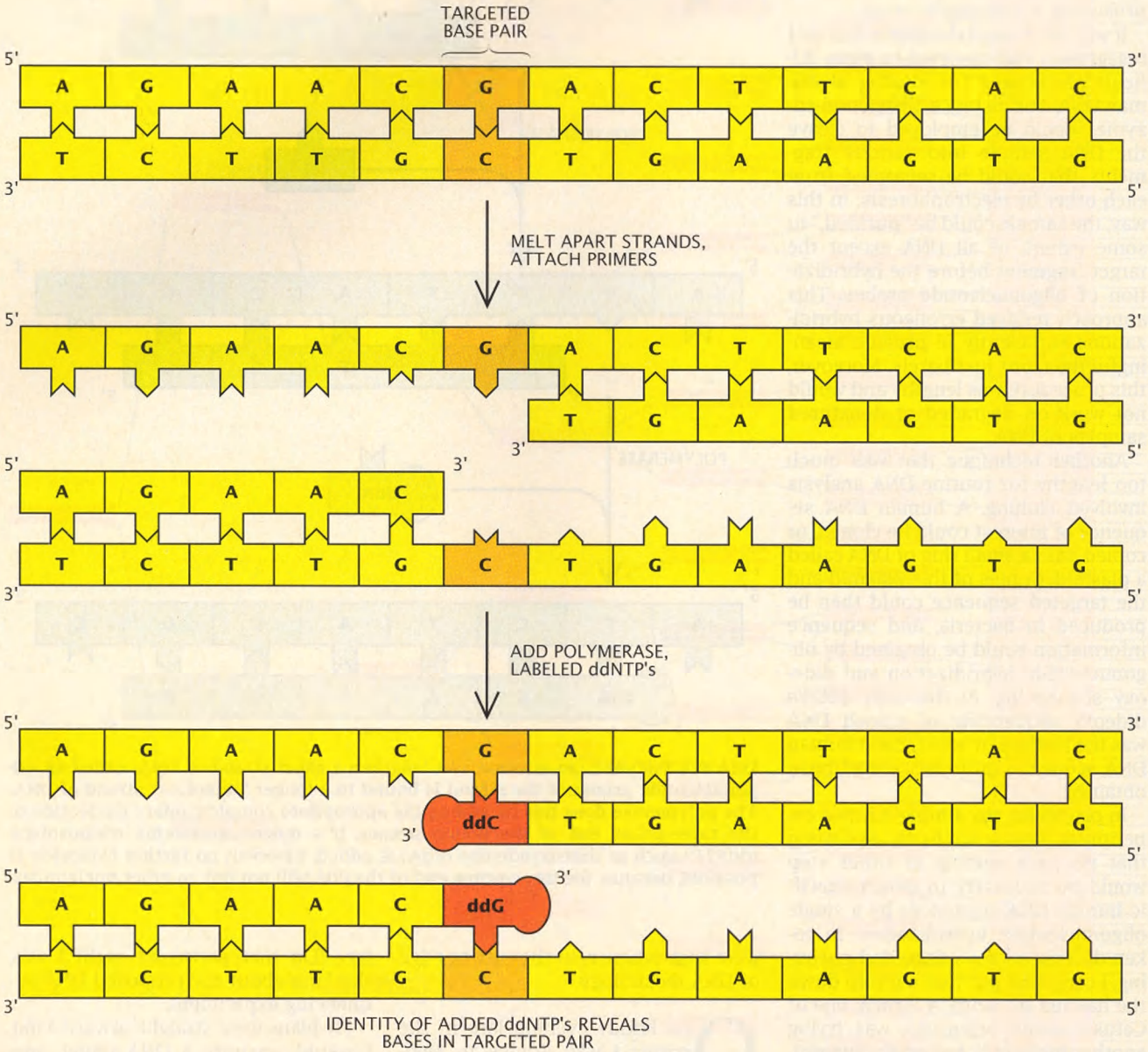
My plans were straightforward. First I would separate a DNA target into single strands by heating it. Then I would hybridize an oligonucleotide to a complementary sequence on one of the strands. I would place portions of this DNA mixture into four different tubes. Each tube would contain all

four types of ddNTP's, but in each tube a different type of ddNTP would be radioactively labeled. Next I would add DNA polymerase, which would extend the hybridized oligonucleotides in each tube by a single ddNTP. By electrophoresis I could separate the extended oligonucleotides from the residual ddNTP's; by identifying which radioactively labeled ddNTP had been incorporated into the oligonucleotide, I could determine the corresponding complementary base in the target strand. Simple.

Around Cloverdale, where California 128 branches northwest from U.S. 101 and winds upward through the coastal range, I decided the determination would be more definitive if, instead of just one oligonucleotide, I used two. The two primers would bracket the targeted base pair I hoped to identify. By making the oligonucleotides of different sizes, I would be able to distinguish them from each other. By directing one oligonucleotide to each strand of the sample DNA target, I could get complementary sequencing information about both strands. The experiment would thereby contain an internal control at no extra inconvenience [*see illustration below*].

Although I did not realize it at that moment, with the two oligonucleotides poised in my mind, their three-prime ends pointing at each other on opposite strands of the gene target, I was on the edge of discovering the polymerase chain reaction. Yet what I most felt on the edge of was the mountain road.

That night the air was saturated with moisture and the scent of flowering buckeye. The reckless white stalks poked from the roadside into the glare of my headlights. I was thinking about the new ponds that



TO DETERMINE THE IDENTITY of a targeted base pair in a piece of DNA, the author hoped to apply a variation on a technique called dideoxy sequencing. First two primers would be bound to the opposing strands in the DNA at sites flanking the targeted pair. DNA polymerase and dideoxynucleotide triphosphates (ddNTP's) would then be added to the mixture, which would allow each of the primers to be extended by only one base. The identity of the added ddNTP bases would reveal what the complementary targeted bases were. The technique could work with only one primer, but the use of two would provide a control for checking the results. Planning this experiment led the author to the polymerase chain reaction.

I was digging on my property, while also hypothesizing about things that might go wrong with my base-sequencing experiment.

From my postdoctoral days in Wolfgang Sadee's laboratory at the University of California at San Francisco, where John Maybaum was devising clinical assays for nucleotides, I remembered that my DNA samples might contain stray traces of nucleotide triphosphates. It would complicate the interpretation of the gel, I figured, if stray nucleotides introduced with the sample added themselves to the three-prime end of the primers before the planned addition of the labeled ddNTP's.

One thought I had was to destroy any loose nucleotide triphosphates in the sample with alkaline phosphatase, a bacterial enzyme. This enzyme would chew the reactive phosphate groups off any nucleotide triphosphates, thereby rendering them inert to a polymerase reaction. Yet I would then somehow have to eliminate the phosphatase from the sample, or else it would also destroy the ddNTP's when I added them. Normally one can deactivate unwanted enzymes by heating them and altering their essential shape; I believed, however, bacterial alkaline phosphatase could refold itself into its original form. I therefore rejected alkaline phosphatase as an answer to the problem.

I was, in fact, mistaken. Much later I learned that alkaline phosphatase can be irreversibly denatured by heating if no zinc is present in the solution. As it turned out, my mistake was extraordinarily fortunate: had I known better, I would have stopped searching for alternatives.

Every mile or so another potential solution arose but fell short. Then, as I began the descent into Anderson Valley, I hit on an idea that appealed to my sense of aesthetics and economy: I would apply the same enzyme, DNA polymerase, twice—first to eliminate the extraneous nucleotide triphosphates from the sample, then to incorporate the labeled ddNTP's.

I reasoned that if there were enough nucleotides in the sample to interfere with the experiment, there would also be enough for the DNA polymerase to act on. By running the sample through a kind of preliminary mock reaction with oligonucleotide primers and polymerase but without ddNTP's, I could easily deplete any nucleotides in the mixture by incorporating them into the extending oligonucleotides. Then, by raising the temperature of the sample, I could separate the extended oligonucleotides from the DNA targets. True, the extended oligonucleotides would still be in the sample; but because there would be far more unextended primers than extended ones in the mixture, the DNA targets would probably hybridize with unextended primers when the mixture cooled. I could then add ddNTP's and more polymerase to perform my sequencing experiment.

Yet some questions still nagged at me. Would the oligonucleotides extended by the mock reaction interfere with the subsequent reactions? What if they had been extended by many bases, instead of just one or two? What if they had been extended enough to create a sequence that included a binding site for the other primer molecule? Surely that would cause trouble....

No, far from it! I was suddenly jolted by a realization: the strands of DNA in the target and the extended oligonucleotides would have the same base sequences. In effect, the mock reaction would have doubled the number of DNA targets in the sample!

Suddenly, for me, the fragrance of the flowering buckeye dropped off exponentially.

Under other circumstances, I might not have recognized the importance of this duplication so quickly. Indeed, the idea of repeating a procedure over and over again might have seemed unacceptably dreary. I had been spending a lot of time writing computer programs, however, and had become familiar with reiterative loops—procedures in which a mathematical operation is repeatedly applied to the products of earlier iterations. That experience had taught me how powerful reiterative exponential growth processes are. The DNA replication procedure I had imagined would be just such a process.

Excited, I started running powers of two in my head: two, four, eight, 16, 32.... I remembered vaguely that two to the tenth power was about 1,000 and that therefore two to the twentieth was around a million. I stopped the car at a turnout overlooking Anderson Valley. From the glove compartment I pulled a pencil and paper—I needed to check my calculations. Jennifer, my sleepy passenger, objected groggily to the delay and the light, but I exclaimed that I had discovered something fantastic. Nonplussed, she went back to sleep. I confirmed that two to the twentieth power really was over a million and drove on.

About a mile farther down the road I realized something else about the products of the reaction. After a few rounds of extending the primers, dissociating the extension products, rehybridizing new primers and extending them, the length of the exponentially accumulating DNA strands would be fixed because their ends would be sharply defined by the five-prime ends of the oligonucleotide primers. I could replicate larger fragments of the original DNA sample by designing primers that hybridized farther apart on it. The fragments would always be discrete entities of a specified length.

I stopped the car again and started drawing lines of DNA molecules hybridizing and extending, the products of one cycle becoming the templates for the next in a chain reaction.... Jennifer protested again from the edge of sleep. "You're not going to believe this," I crowed. "It's incredible."

She refused to wake up. I proceeded to the cabin without further stops. The deep end of Anderson Valley is where the redwoods start and where the "ne'er-do-wells" have always lived. My discovery made me feel as though I was about to break out of that old valley tradition. It was difficult for me to sleep that night with deoxyribonuclear bombs exploding in my brain.

Yet in the morning I was too tired not to believe that someone, somewhere, must have tried this idea already. Thousands of investigators had, for various reasons, extended single oligonucleotides with polymerases; surely someone would have noticed the possibility of a polymerase chain reaction. But if it had worked, I was sure I would have heard about it: people would have been using it all the time to amplify, or multiply, DNA fragments.

Back at Cetus on Monday I asked one of the librarians, George McGregor, to run a literature search on DNA polymerase. Nothing relevant to amplification turned up. For the next few weeks I described the idea to anyone who would listen. No one had heard of its ever being tried; no one saw any good reason why it would not work; and yet no one was particularly enthusiastic about it. In the past, people had generally thought my ideas about DNA were off the wall, and sometimes after a few days I had agreed with them. But this time I knew I was on to something.

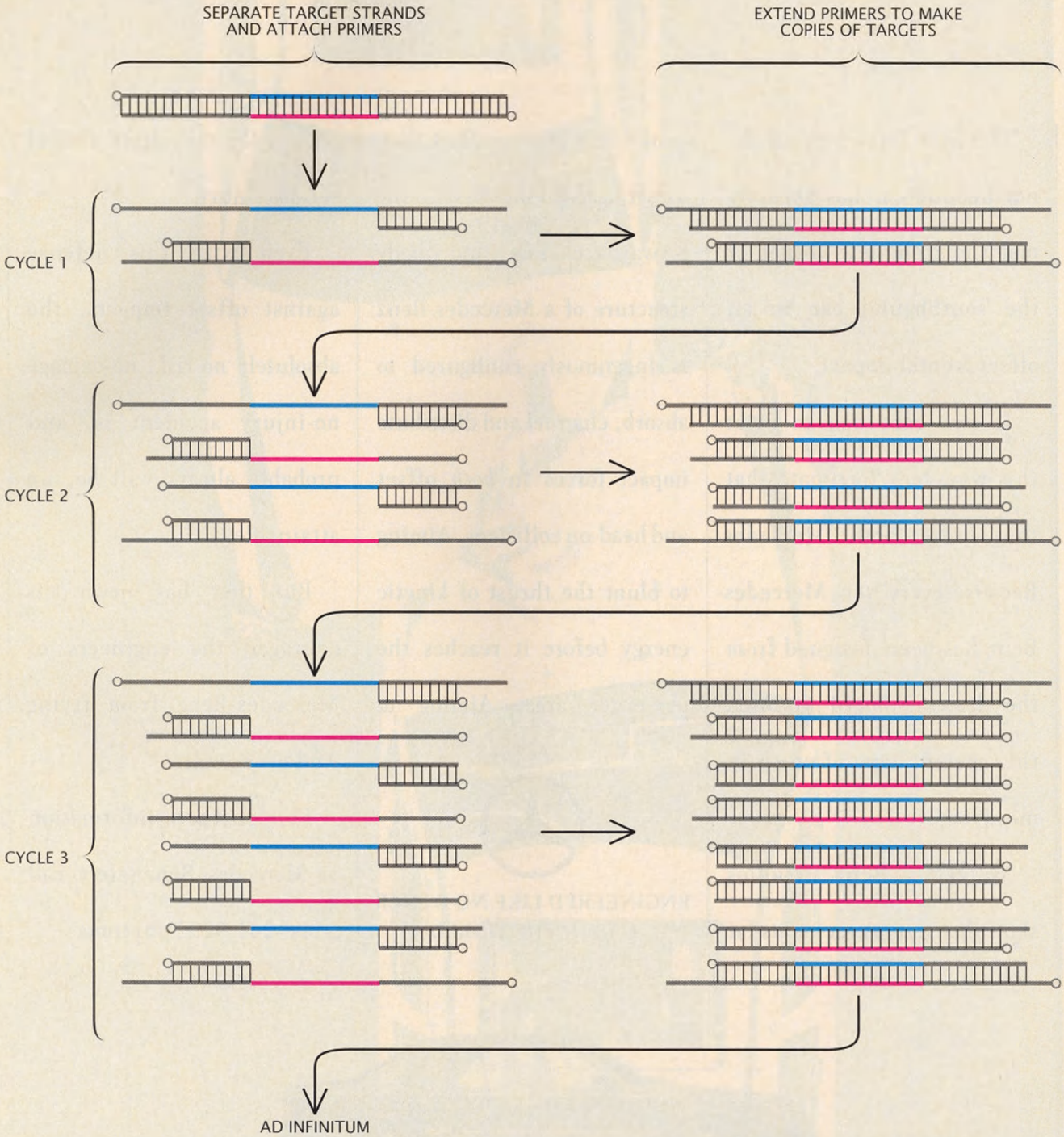
Years ago, before biotechnology—when being a genetic engineer meant that you, your dad and his dad all drove trains—our building at Cetus had been owned by the Shell Development Company. Our laboratory space,

whose rear windows looked grandly out on the Berkeley hills, had given birth to the "No-Pest Strip." It did not escape my notice that the PCR might someday travel as far as its sibling invention, that distinctively scented piece of yellow plastic.

Months passed as I prepared for my first experiment to verify whether the PCR would work. I had to make many educated guesses about what buffer solutions to use, what the relative and absolute concentrations of the reactants should be, how much to heat and cool the mixtures, how long the mixtures should run and so on. Some of Kornberg's early papers on DNA polymerase helped. To run the experiment, I selected a 25-base-pair target fragment of a plasmid and two oligonucleotide primers that were 11 and 13 bases long, respectively.

When everything was ready, I ran my favorite kind of experiment: one involving a single test tube and producing a yes or no answer. Would the PCR amplify the DNA sequence I had selected? The answer was yes.

Walking out of the lab fairly late in the evening, I noticed that Albert Halluin, the patent attorney for Cetus, was still in his office. I told him that I had invented something and de-



POLYMERASE CHAIN REACTION is a cyclic process; with each cycle, the number of DNA targets doubles. The strands in each targeted DNA duplex are separated by heating and then cooled to allow primers to bind to them. Next DNA polymerases extend the primers by adding nucleotides to them. In this way, duplicates of the original DNA-strand targets are produced.

scribed the PCR. Al was the first person, out of maybe a hundred to whom I had explained it, who agreed that it was significant. He wanted to see the autoradiogram showing the experimental data right away; it was still wet.

Some people are not impressed by one-tube experiments, but Al was not noticeably skeptical. Patent attorneys, after all, have a vested interest in inventions. He had followed my explanation of the process in his office and agreed that it made sense. Now in the lab he was even a little excited and suggested that I get to work on the experiment and write a patent disclosure. As he left, he congratulated me.

For the next few months I continued to study and refine the PCR with the help of Fred A. Faloona, a young mathematics wizard whom I had met through my daughter. Fred had helped me with the first PCR experiment by cycling the DNA mixture—in fact, that had been his very first biochemistry experiment, and he and I celebrated on the night of its success with a few beers.

In the following months we confirmed that the PCR would work on larger and larger fragments of plasmid DNA. Eventually we obtained some human DNA from Henry Erlich's laboratory and produced evidence for the amplification of a fragment from a single-copy gene.

Today many of the initial hitches or inefficiencies of the PCR have been worked out. Several slightly different protocols are now in use. I usually recommend that the DNA samples be cycled between temperatures of about 98 degrees Celsius, just below boiling, and about 60 degrees C. These cycles can be as short as one or two minutes; during each cycle the number of DNA target molecules doubles. The primers are usually from 20 to 30 bases long. One of the most important improvements in the process is the use of a particular DNA polymerase originally extracted from the bacterium *Thermus aquaticus*, which lives in hot springs. The polymerase we had originally used was easily destroyed by heat; consequently, more had to be added during each cycle of the reaction. The DNA polymerase of *Thermus aquaticus*, however, is stable and active at high temperatures, which means that it only needs to be added at the beginning of the reaction. This high-temperature polymerase is now produced conveniently by genetically engineered bacteria.

The virtually unlimited amplification of DNA by the PCR was too unprecedented to be accepted readily. No one was prepared for a process that provided all the DNA one could want. The reaction seemed self-evident to Fred and to me because it was our toy. For most people, it took some getting used to.

MACHINE that performs the polymerase chain reaction is shown being loaded with samples of DNA. Such devices are rapidly becoming common fixtures in laboratories.

In the spring of 1984, while working on the patent, I presented a poster describing the PCR at the annual Cetus Scientific Meeting. These meetings were always fun, because Cetus had some first-rate scientific advisers, and I was looking forward to talking with them about my invention.

Yet nobody seemed to be interested in my poster, and I felt increasingly anxious. People would glance at it and keep walking. Finally, I noticed Joshua Lederberg, president of the Rockefeller University, nearby, and I snared him into looking at my results. Josh looked the poster over carefully and then turned his enormous head, the Nobel-laureated head, the head that had deduced in 1946 that bacteria could have sexual intercourse. "Does it work?" He seemed amused.

Pleased, I confirmed that it did, and we talked for a long time. At one point he mentioned that about 20 years previously, after Kornberg had discovered DNA polymerase, the two of them had considered the notion that the enzyme could somehow be harnessed to make large quantities of DNA. They had not figured out exactly how to do it, however. I reminded him that oligonucleotides were not readily available at that time and that there was hardly any DNA sequence information either.

But he looked back at my poster with an expression that I have almost come to expect. I think that Josh, after seeing the utter simplicity of the PCR, was perhaps the first person to feel what is now an almost universal first response to it among molecular biologists and other DNA workers: "Why didn't I think of that?" And nobody really knows why; surely I don't. I just ran into it one night.

FURTHER READING

SPECIFIC ENZYMATIC AMPLICATION OF DNA IN VITRO: THE POLYMERASE CHAIN REACTION. Kary Mullis, Fred Faloona, Stephen Scharf, Randall Saiki, Glenn Horn and Henry Erlich in *Cold Spring Harbor Symposia on Quantitative Biology*, Vol. 51, No. 1, pages 263–273; 1986.

SPECIFIC SYNTHESIS OF DNA IN VITRO VIA A POLYMERASE-CATALYZED CHAIN REACTION. Kary B. Mullis and Fred A. Faloona in *Methods in Enzymology*, Vol. 155, Part F, pages 335–350; 1987.

AMPLIFICATION OF HUMAN MINISATELLITES BY THE POLYMERASE CHAIN REACTION: TOWARDS DNA FINGERPRINTING OF SINGLE CELLS. Alec J. Jeffreys, Victoria Wilson, Rita Neumann and John Keyte in *Nucleic Acids Research*, Vol. 16, pages 10953–10971; 1988.

DNA SEQUENCING WITH THERMUS AQUATICUS DNA POLYMERASE AND DIRECT SEQUENCING OF POLYMERASE CHAIN REACTION-AMPLIFIED DNA. M. A. Innis, K. B. Myambo, D. H. Gelfand and Mary Ann D. Brow in *Proceedings of the National Academy of Sciences*, Vol. 85, No. 24, pages 9436–9440; December, 1988.

THE POLYMERASE CHAIN REACTION. T. J. White, Norman Arnheim and H. A. Erlich in *Trends in Genetics*, Vol. 5, No. 6, pages 185–188; June, 1989.