end of the chief "secondary structures" within the larger architecture of proteins: alpha helixes and the folds of the so-called beta pleated sheets. By equipping artificial proteins with these sequences, Richardson and her colleagues were able to mimic the overall shape of many natural proteins: a compact globule consisting of a bundle of helixes.

This approach hasn't stopped being fruitful. Du Pont chemist William DeGrado says he's taken it a step further by identifying short sequences—even single amino acids—that can help constrain a protein's configuration by ruling out alternatives. Since an amino acid sequence can usually fold into several different structures, both natural and artificial sequences must include safeguards to prevent then from snapping into undesirable configurations. "It's as if you are giving someone instructions to go from L.A. to New York and you have to tell them not to get off at every exit before the one you want." For example, when DeGrado tried to design a double-helical structure, he found he had to alter the sequence to prevent one helix from doubling again, forming a triple helix.

Princeton's Hecht, meanwhile, suspected that there might be more general rules for protein structure that don't depend on the presence or absence of specific amino acids. He and others had noted an underlying pattern, for example, in the sequence of natural proteins that have a quadruple coil structure known as a four-helix bundle: Sets of hydrophilic (water attracting) and hydrophobic (water repelling) amino acids alternate at intervals along the sequence. The intervals are just the right length to place the hydrophobic amino acids on the inside of each coil and the hydrophilic ones on the outside.

Hecht and his colleagues mimicked this pattern in a series of synthetic proteins, choosing amino acids that showed the alternating hydrophilic/hydrophobic pattern, but were otherwise chosen at random. The researchers then analyzed the test proteins by measuring their optical properties—a rough indicator of structure. To Hecht's surprise, 60% of the several hundred proteins had folded into four-helix bundles (*Science*, 10 December 1993, p. 1680). "No one thought such semi-random sequences would fold into stable structures," he says. He believes that result implies that overall protein structure doesn't always depend on the presence or absence of specific amino acids. Instead, much more general chemical properties may be capable of determining form.

Interior design. Still, the rules deciphered so far aren't nearly enough to specify the full architecture of a protein, as Hecht, DeGrado, and Richardson have found. The interior of a protein, crowded with branching offshoots known as side chains, is still largely terra incognita. In natural proteins, these chains, whose position is critical to functions such as catalyzing chemical reactions, are locked, jigsaw-puzzle fashion, in just the right pattern. But in many synthetic proteins, Hecht says, the side chains are "flopping in the breeze."

DeGrado is now concentrating on learning how natural proteins pin down their side chains—and how he can do the same, whether by nature's means or his own. In 1992 he showed that added metal atoms, each of which can bind to two or three side chains, can stabilize the interior of a synthetic protein. But while that finding could aid researchers trying to design artificial molecules, it doesn't add much to the understanding of natural proteins, most of which don't rely on metals for stability. More recently DeGrado has experimented with artificial sequences in which the amino acids alternate in size, which should allow them to pack in a stable, orderly array. First published in 1991 in the *Journal of the American Chemical Society*, the approach has since yielded structures with four helixes that, says DeGrado, have the interior order of true proteins.

Researchers eager to design functional proteins aren't waiting for Hecht, DeGrado, and their colleagues to decipher all the rules of protein folding. Leslie Dutton of the University of Pennsylvania, for example, recently applied some of the known rules to devise the first artificial protein that has a stable interior and is also designed to carry out a specified function. Dutton simplifies the problem of controlling a synthetic protein's structure by streamlining it, reducing the molecule to a kind of scale model of the real thing. "We identify what we think are the key elements in a natural protein, and then synthesize these in a test tube without the biological baggage," he says.

In Dutton's simplified protein, the problem of the floppy interior solved itself. As a scaffolding for his protein, he used the multiple helix structure chemists are now adept at designing, then attached four "functional" groups called hemes. The bulky heme groups, he found, propped up the artificial protein's floppy interior, yielding a molecule that had the rigidity of a natural protein, along with a potentially useful property: the ability to transfer electrons, a function crucial in photosynthesis and cellular respiration.

Petsko, too, isn't waiting for the protein-folding rulebook to be completed. To speed up the search for "rational" drugs targeted to disease proteins, he is pushing ahead with traditional x-ray crystallography, cataloguing the structures of natural proteins. But once biologists can reliably predict protein structure from amino acid sequences, his difficult and time-consuming work will be a thing of the past. And Petsko says he's looking forward to that day. "I would love to be put out of business," he says.

–Faye Flam

MOLECULAR BIOLOGY

'Long PCR' Leaps Into Larger DNA Sequences

On Christmas Eve 1992, plant molecular biologist Wayne Barnes showed his son one of Barnes' own new toys: a video machine that displays sharp images of lab gels, which contain molecules separated by size. Barnes, of Washington University School of Medicine in St. Louis, slid in a gel from a recent experiment using the polymerase chain reaction (PCR) technique. PCR amplifies genetic material, making it possible to study specific genetic sequences, and Barnes was pushing the technique to copy longer pieces of DNA than it had ever copied before. When he saw the gel, and the size of the DNA molecule on it, Barnes realized he had succeeded. "I said, 'Boy, this is a fairly big result,'" recalls Barnes.

And, he adds, "big had two meanings." The first meaning of big pertains to the size of the DNA molecules that Barnes can now copy. PCR, invented in 1983, has had a revolutionary impact on molecular biology because it offers an easy way to copy large quantities of DNA, something that was previously a laborious task. Yet until now the technique couldn't reliably copy DNA molecules longer than about 5000 nucleotide bases, putting many molecules of great biological interest out of reach. In the 15 March issue of the *Proceedings of the National Academy of Sciences* (*PNAS*), however, Barnes reports that by optimizing the chemical conditions of the reaction he is able to amplify stretches of DNA seven times as large: 35 kilobases (kb).

The second meaning of big is the impact that large segments could have. "If you had a project that required 10kb before, it was undoable," says Barnes. "Now it's dead easy." In addition to making work easier for scientists studying the human genome, by giving them longer contiguous DNA sequences to analyze, researchers say this "long PCR," as Barnes dubs it, may help them diagnose genetic diseases or copy and study HIV.

Barnes' big leap is only one of several innovations that have begun to make work on PCR seem like molecular biology's version of the broad jump. Scientists at Roche Molecular Systems in Alameda, California, who were trying to make longer PCR products themselves, began collaborating with Barnes last year and have submitted a separate paper to *PNAS* about their progress.

Another approach, developed by Elise Rose at Perkin-Elmer, has produced what she calls "long-distance" PCR, and she has a review paper on the topic in press at *PCR Methods and Applications*.

PCR jocks are both intrigued by and tentative about the findings leaping out of these labs. "I want to stress that I don't know whether these methods will work," says Eric Lander, head of the Whitehead/MIT Center for Genome Research. But if long PCR does pan out, Lander says, "it could be spectacularly useful. PCR at short lengths transformed our lives. PCR at long lengths could transform our lives again." Richard Gibbs, a molecular geneticist at Baylor College of Medicine, who is working on mapping the X chromosome, also is cautiously enthusiastic. "I think it may change many of the things we do," says Gibbs.

SOURCE: W. BARNES ILLUSTRATION: K. SUTLIFF

PCR first became popular because it's essentially a technique for taking over nature's machine for replicating genetic material and shifting it into overdrive. That technique requires several steps. First, DNA is heated to separate the two complementary strands of the double helix. Each strand then becomes a template for making a mirror-image copy of itself. To make that mirror image, short DNA pieces called "primers" attach to the end of the template. DNA polymerase, an enzyme that builds new DNA strands, starts at the attachment point and extends the primers nucleotide by nucleotide until the new mirror-image copy is made. New cycles of the reaction exponentially amplify the DNA.

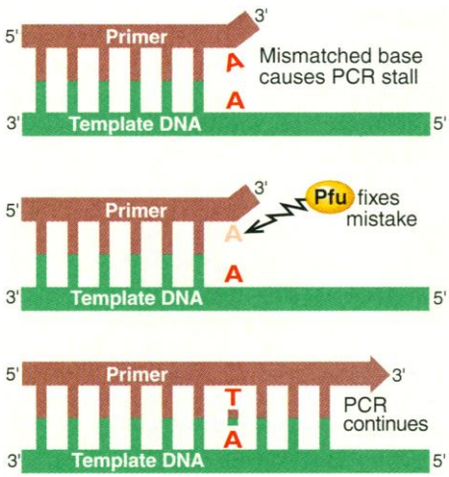
Barnes' idea for improving this process struck him as he tried to use standard PCR to engineer plants to express a protein from *Bacillus thuringiensis* (*Bt*), a bacterium that kills beetles and moths. Barnes knew the sequence of a 2 kb *Bt* gene and wanted to make his own clone, so he could insert it into a plant genome. A DNA segment of that length was well within the range of standard PCR. "The whole process should have been done in 2 or 3 days if it worked," says Barnes. "I'm lucky it didn't, because I wouldn't have found the trick" for long PCR.

It turned out that Barnes had used a primer for the *Bt* gene that had the wrong DNA nucleotide on one end. There are strict rules for which nucleotide bases can pair with each other on matching DNA strands. The template strand had an adenine (A), which can only pair with a thymine (T). The primer, however, also had an A. When the A on the primer met the A on the template, the mismatch apparently caused the PCR engine to seize.

In this experiment, Barnes had used a DNA polymerase called Klentaq1. To solve the problem of the mismatch, he tried a different polymerase, *Pfu*, one that specifically chews up nucleotides on the ends of DNA strands. It, too, failed. Concerned that the *Pfu* was chewing up too much of the primer, Barnes hit on the idea of combining Klentaq1 with very low concentrations of *Pfu*. It worked: The amplification kicked into gear. "The *Pfu* was needed to clean off the mismatch and allow the Klentaq1 to charge ahead," theorizes Barnes.

Then Barnes began wondering: If a mismatch on the primer end stopped PCR, wouldn't a mismatch accidently created by the polymerase, which extends the primer, have the same effect further down the line? That would limit PCR to short strands. Barnes tried his two-punch polymerase idea and soon, with a little jiggling of other conditions, such as the time of each amplification cycle, he was up to 28 kb.

At about the same time, over at Roche,



Back on track. A pairing mismatch may derail DNA polymerase as it tries to copy a template during PCR. A second polymerase removes the mismatch, allowing the reaction to proceed.

Suzanne Cheng, Russell Higuchi, and Carita Fockler were also making longer and longer PCR products by adjusting their chemical conditions. The Roche scientists believed a major limiting factor to long product was that the PCR process itself was damaging the DNA. The high temperatures used to separate the DNA strands might cause breaks in the template DNA and prevent the polymerase from adding nucleotides. The pH also might be too low, breaking bonds that hold nucleotides on the new strand.

To remedy the first problem, the Roche team added co-solvents, such as glycerol, which helped the DNA strands to separate at lower temperatures. They also maintained the pH in the reaction vessel by using buffering agents.

With these and a few other modifications, the Roche group managed to amplify a λ bacteriophage DNA template of 26kb and even amplified 12kb from more complex human genomic material. Roche researchers began meeting with Barnes in early 1993, and soon realized they stood to gain by trading their secrets. The shared information allowed Barnes to produce high, clean yields of 35kb λ bacteriophage DNA.The Roche team, meanwhile, adopted a double polymerase strategy similar to Barnes' and now claim to have taken λ out to 42kb. Working with the human β-globin gene cluster, the Roche group says they now have evidence that they have amplified a 22kb fragment. "[Barnes'] idea doubled the distance we could go," says Cheng.

Rose, at Perkin-Elmer, has a different idea about the limitations to long-distance PCR, and a different way around them. She believes the key problem with going long distances is that polymerases only work for a certain amount of time before falling off. The trick, then, is to find polymerases that stay on the template for relatively longer periods, copy nucleotides more quickly, or both. Using polymerases such as *TTh* and *Vent*, Rose says she routinely goes out as far as 20-25 kb with human DNA. As for the A-A mismatch hypothesis, Rose says, "I don't see strong evidence for that."

No matter who's right about the specific mechanism, any of these long PCR systems could open up a more spacious world for geneticists if they work consistently. One of the things that could change is mapping the human genome. Maps of the genome already exist, but they are on the scale of states or cities rather than city blocks or individual houses. Locating a specific gene on this map is similar to looking for a street address. This typically requires cloning DNA in yeast or bacteria, and then hunting for artifacts and errors in the sequences. Long PCR avoids these pitfalls by directly making longer stretches of DNA available for analysis. This should make it easier for researchers to spot the contiguous sequence of nucleotides making up their current object of interest.

If all goes well, Roche plans to market a long PCR kit in the next few months, and a myriad of applications promise to follow quickly. "Academics are going to get a hold of it and fly with it," predicts Perkin-Elmer's Rose, whose company makes PCR machines. Already, Barnes says he has used long PCR to amplify the entire HIV genome, which might make it much simpler to sequence the virus and study variation, a key issue in the search for an AIDS vaccine. Long PCR also is being experimented with as a diagnostic, to see whether it can help distinguish a healthy patients' DNA from a patient who has extensive deletions or insertions in a specific gene.

And Rose believes that long-distance PCR has yet to approach the edge of the envelope. "If you can simulate what goes on in a cell, you should be able to go out several hundred kb," she theorizes, noting this is the natural limit. Big dreams, in this case, don't start small.

–Jon Cohen