# Making the Cut

## Scientists exploit proteins that perform surgery upon themselves

By JOHN TRAVIS

In the lungs of a person with tuberculosis, immune cells called macrophages do battle with the hardy bacterium responsible for the disease. Macrophages bomb the microbe, *Mycobacterium tuberculosis,* with oxygen-containing compounds that can damage DNA. *M. tuberculosis* fights back, but not by defusing these highly reactive compounds, as many other infectious bacteria do.

Instead, it depends upon its skill at repairing DNA to survive. In what biologists call the SOS response, the bacterium activates a protein called RecA. "It's a trigger that turns on the whole cascade of gene expression that produces the DNA-repair machinery," says Henry Paulus of the Boston Biomedical Research Institute.

Before it fires up the SOS response, however, RecA must accomplish a molecular sleight of hand that only a few proteins can perform. A small portion of RecA, dubbed the intein, quickly cuts itself out of the protein and melds the two remaining fragments together to form the final, active molecule.

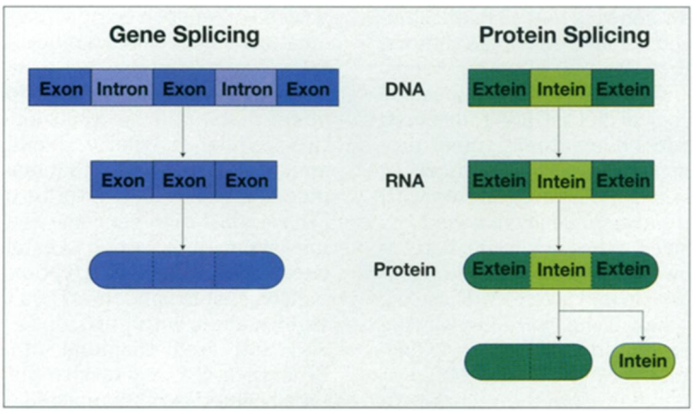
Since 1990, biologists have observed this bizarre phenomenon, called protein splicing, in about 100 proteins in some 30 single-celled organisms throughout nature. Despite this rarity, scientists have learned a lot over the past decade about the mechanics of intein action. They also have learned how to put these proteins within proteins to work.

Investigators now use inteins to purify proteins or bend them into loops, for example. Inteins may one day even help scientists make artificial spider silk or control DNA given to patients during gene therapy.

"There are so many cool things that people are doing," says Francine B. Perler of New England Biolabs in Beverly, Mass., a pioneer in protein splicing.

While biotechnology applications of inteins have blossomed, biologists have made little progress in revealing the roles that inteins naturally play inside a cell—or even if they have a cellular function at all. Scientists understand the role of RecA and realize that it has an intein, but they have no idea whether cells have developed ways to take advantage of the presence of this intein or any other.

"Right now, their biological function is just speculation," says Paulus. "The time has come that we address function. No one has seriously done that."




*In gene splicing, a cell removes DNA segments called introns before using the rest of the gene, the exons, to make a protein. In protein splicing, a protein fragment, the intein, removes itself from the full protein and fuses the remaining parts, the exteins.*

Determining the purpose of inteins might have a medical payoff. Compounds that inhibit RecA's protein-splicing ability, suggests Paulus, could offer a new way of treating tuberculosis.

In the beginning, genetics was simple. Scientists thought that cells read a continuous DNA sequence, the gene, and translate its encoded information into the amino acids that form a protein. Then, biologists discovered that many genes are a mix of protein-coding DNA sequences, which are called exons, and noncoding sequences, called introns. Before these genes form any protein, cells cut out all the introns and splice together the remaining exons.

Almost a decade ago, investigators began to see hints that proteins could perform a comparable feat. In 1990, two research teams independently reported that a yeast protein seemed to cut out an interior section and fuse the two flanking pieces. Yet this reaction proceeded so quickly that scientists couldn't actually identify the full-length, original protein.

A year or two later, Perler's group was working with a protein from a microbe found in a deep-sea vent. Like the yeast protein, this one appeared to carve pieces out of itself and rejoin the leftover parts. By placing the gene of the deep-sea protein into a different bacterium, Perler and her colleagues slowed this transformation and actually isolated the precursor protein.

Since the researchers reported this proof that protein splicing happens, investigators have worked to describe the steps by which an intein cleaves the chemical bonds that hold it within the precursor protein and then seamlessly joins the protein fragments from which it has escaped. The intein unites these fragments, dubbed exteins, with the same natural bond between amino acids that cells employ when they build a protein from scratch.

Once biochemists realized the skills of inteins, they quickly began to dream up uses for the talented proteins. New England Biolabs now markets an intein-based protein-purification system. Scientists needing to collect large quantities of a particular protein can take the gene for that protein and fuse it to the DNA sequence encoding a modified intein. They then slip this fused gene into bacteria, which produce the protein-intein hybrid.

These hybrids are easy to separate from other bacterial molecules because the company engineered the intein to stick to a protein called chitin. After using chitin-coated beads to pluck the hybrids out of a mix of proteins, scientists chemically activate the intein to cleave its bond with the desired protein. Voilà, the intein stays stuck to the chitin beads and investigators have the protein they want.

Last year, Thomas C. Evans Jr., Jack Benner, and Ming-Qun Xu of New England Biolabs described a way of using inteins to make proteins that are normally toxic to bacteria and other cells. The investigators engineered two strains of bacteria to make different halves of the cytotoxic protein, each one attached to half of an intein. When they mix these engineered proteins in a test tube, the intein fragments find their opposite half, reassemble, cut themselves out, and splice together the whole cytotoxic protein.

Biochemists have also modified inteins

so that when they hack themselves from a protein, they leave the exposed protein ends susceptible to forming a bond with another protein. Thus, the intein creates a bit of "molecular Velcro" on the protein as it departs, says Tom W. Muir of Rockefeller University in New York City.

Muir's group and Xu's have used this tactic to make proteins with sticky ends. If a protein's structure is flexible, those ends can join to each other and transform the protein into a loop, the teams report in the June 16 JOURNAL OF THE AMERICAN CHEMICAL SOCIETY and the June 25 JOURNAL OF BIOLOGICAL CHEMISTRY, respectively.

Why do scientists play such molecular games? For some proteins, circularizing may stabilize them and alter their activity, says Evans, noting that his company looks for enzymes that are more active, heat-tolerant, or acid-resistant than the natural versions.

Evans and his colleagues are also hoping that inteins and their ability to create proteins with sticky ends will help them combine proteins into long chains. For example, the scientists have begun to examine whether they can use inteins to link the proteins in spider silk, which is prized for its strength and flexibility (SN: 3/9/96, p. 152).

Muir's group has also used inteins to splice a small fluorescent region into the middle of a protein. The brightness of the resulting protein's glow depends on its activity, a feature that should help scientists study the protein's roles in cells.

In addition, inteins may aid researchers trying to determine the three-dimensional structures of proteins. The two main techniques used today for such a purpose are X-ray crystallography, in which X-ray beams are shot through a crystallized protein, and nuclear magnetic resonance (NMR) spectroscopy, which uses a strong magnetic field to map the nuclei of atoms. Many scientists prefer NMR spectroscopy since it probes proteins in a watery solution, their natural environment, but the method can't decipher the structure of large proteins.

Muir and his colleagues believe inteins can help overcome the NMR size barrier. They fuse intein genes to a section of the gene for a protein they want to study and use that DNA to make a protein fragment containing carbon and nitrogen isotopes that are more easily detected by NMR than the normal ones. The investigators then use the intein to splice that fragment to the rest of the unhighlighted protein.

"It's possible to prepare proteins in which only discrete parts of the molecule are labeled with nuclei that are sensitive to NMR spectroscopy," says Muir. By labeling different parts of the protein, the investigators can resolve the shapes of the individual regions and then merge them into a complete protein structure.

Overall, the protein-splicing ability of inteins has opened a new world of possibilities for biochemists trying to synthesize molecules. Previously, through solely chemical means, researchers were able to string together up to 150 amino acids. "You can make really small proteins using the totally chemical approach, but to do larger [proteins] is impossible," says Muir.

Biochemists have been using bacteria to make longer amino acid chains, but they can't handle the largest of mammalian proteins. Inteins can make that task feasible. The scientists isolate the gene for a large protein, chop it into smaller pieces, and fuse those pieces with intein genes. Bacteria with these fused genes can then make large amounts of intein-bound protein fragments. These fragments are later mixed in a test tube where the inteins reassemble whole copies of the molecule.

Biochemists have found myriad ways to exploit inteins, but it's still not clear whether they have any purpose in nature. Around 70 percent of the known inteins occur in proteins that participate in synthesis, repair, or breakdown of DNA, yet this seemingly nonrandom distribution may not prove significant.

Many of the first introns discovered were in genes underlying DNA metabolism, but the introns don't discernibly influence the process, says Marlene Belfort of the Wadsworth Center of the New York State Department of Health in Albany.

The parallels between introns and inteins extend even further. Introns often encode a small enzyme, a so-called homing endonuclease, that can cut DNA strands, and they use it to move their DNA throughout the genome. Likewise, many inteins contain an endonuclease section in addition to a protein-splicing region. Scientists normally eliminate this DNA cutter from the inteins they use in biotechnology applications. Inteins with an intact homing endonuclease can apparently slice the DNA that encodes them out of a gene and transplant that DNA elsewhere.

What advantage might such enzymes give introns and inteins? Biologists frequently describe introns as "selfish" genetic elements. They seem to have no role other than to move about the genome. Belfort believes that the DNA encoding inteins is similarly selfish.

Since an intein can repair the protein of any gene its DNA interrupts, intein genes can apparently hop around a cell's genome without altering proteins as other mutations would. "We have not been able to demonstrate any function" for inteins, says Belfort.

Paulus, however, argues that scientists need to keep looking. He's intrigued that not all bacterial RecA proteins have inteins. Only *M. tuberculosis* and another mycobacterium, which causes leprosy, do. Since closely related, but harmless, members of the mycobacteria family have uninterrupted RecA proteins, Paulus speculates that the inteins help the two microbes cause disease. Perhaps the bacteria can control when the intein functions and therefore regulate the ability of RecA to turn on DNA repair, he says.

"What we still have to discover is how protein splicing is controlled. That's a completely unexplored area," says Paulus.

By inserting the gene for the tuberculosis bacterium's RecA into *Escherichia coli*, a bacterium much more easily grown in the laboratory, Paulus plans to identify inhibitors of the intein's protein-splicing talent. He would then test any such inhibitors as possible drugs against *M. tuberculosis*.

There's at least one case, notes Paulus, in which a microbe depends upon an intein for its survival. Last year, Xiang-Qin Liu of Dalhousie University in Halifax, Nova Scotia, and his colleagues reported the discovery of a "split intein" while studying a DNA polymerase, the enzyme that replicates DNA, in a strain of bacteria of the genus *Synechocystis*. The biologists realized that the enzyme's gene was in two pieces. The gene segments had many genes between them on the bacterium's circular chromosome. How could the bacteria make a functional polymerase?

Liu's group found that each polymerase gene segment ended in half of an intein gene. The microbe apparently synthesizes each polymerase half with its intein fragment. When the fragments bump into each other in the cell, the intein assembles itself and splices together the whole enzyme. "You don't have to have a coherent gene to make a coherent protein," notes Liu.

While he has yet to find any evidence that the bacterium controls the actions of this split intein, Liu agrees with Paulus that cells may have learned to exploit inteins for regulating protein activity.

"Once something exists in a cell, there's always the opportunity for the element to be recruited into some biological function," he says. Protein splicing "introduces an extra step into making a protein. That provides another opportunity to regulate the function of those proteins," Liu says.

Whether or not cells regulate inteins, scientists are trying to develop their own controllable intein that they can incorporate into proteins. Such an insert might enable them to add a gene to organisms—even people—that would produce an active protein only when the scientists provide a compound that frees the intein.

There seems to be no end to what researchers would like to do with inteins. "They're interesting from an evolutionary standpoint, from a mechanistic standpoint, and from an applied standpoint," says Belfort. "I think there's going to be a Nobel prize in the field, although I'm not sure who's going to get it." □