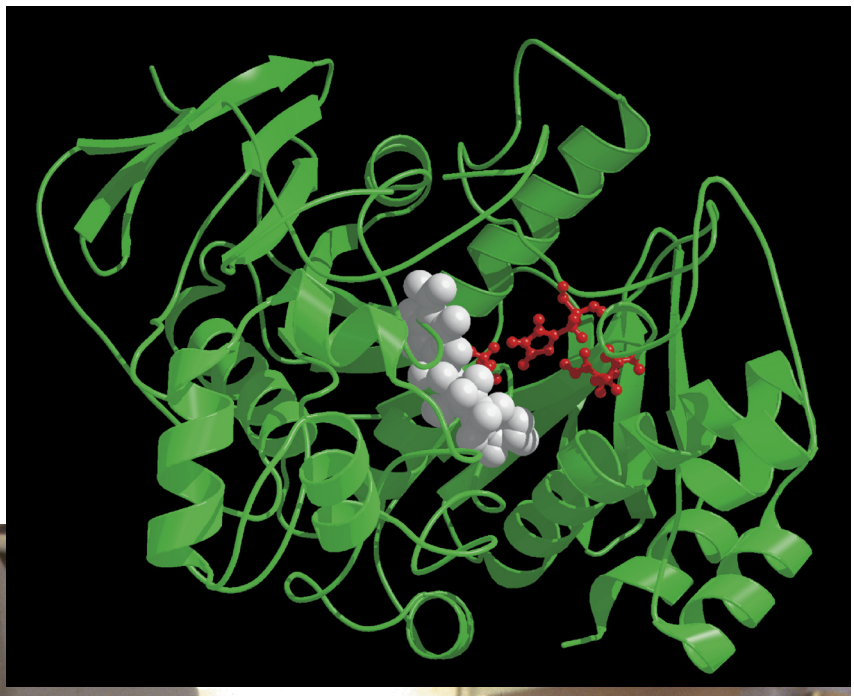
We jump a natural enzyme through a new hoop, and accumulate mutations that help it jump higher.



Unnatural Selection: Molecular Sex for Fun and Profit

by Frances H. Arnold

Top, left: Enzymes are molecular machines whose intricate shapes allow them to function. Here the protein backbone (green) cuddles the reacting molecule (gray) while holding a few amino acids (red) in just the right position to catalyze a reaction within the gray molecule.

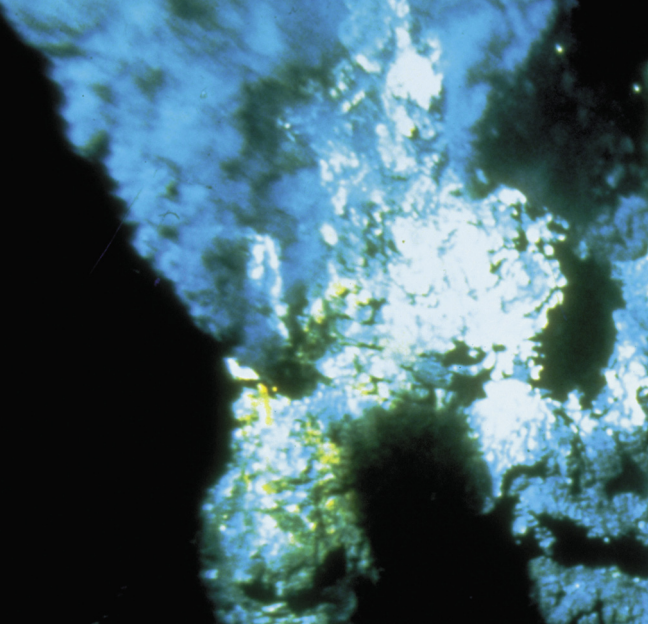
Bottom: Proteins are really big molecules, but this is probably not the best way to modify them, as Mark Tomusiak (MS '91) and Ed Naranjo (BS '89) discover.

Proteins are nature's molecular machines. They're responsible for virtually all the interesting things that biological systems do. Enzymes, for example, are the ultimate chemists—they catalyze or direct all of life's reactions, and they're so remarkably specific that hundreds of reactions can proceed simultaneously inside a tiny cell. (One trillion *E. coli* bacteria will fit in a cubic centimeter.) This ability to synthesize complex materials at room temperature and pressure, in water, without waste products, rightfully earns enzymes the envy of synthetic chemists and the admiration of chemical engineers. One of the 21st century's challenges will be improving the world's standard of living without destroying our environment. That's going to be hard to do, but one way might be to recruit these highly efficient biological systems to work for us. My vision is of a biotechnology-based chemicals industry that makes no messes to clean up.

Now, the chemical engineer who actually has to implement this vision is constantly stymied by the fact that all these enzymes evolved over billions of years to perform very specific biological functions within the context of a living system. But the demands of industry are very different. You take an enzyme out of its natural context, and you find that many of its features are completely incompatible with cost-effective manufacturing. For example, imagine the chemical engineer's frustration with a catalyst that turns itself off the minute it produces a little bit of product. However, this control is vital to the cell, which carefully regulates its metabolism through such feedback loops. Industry wants enzymes that are highly stable (proteins don't take heat well—think of a hard-boiled egg), that can function in organic solvents (because many things we want to make aren't soluble in water), and that react with substances nature never even thought of. We'd even like to have molecules that perform reactions nature doesn't use. To do this we have to engineer enzymes at the molecular level—to redesign them for industrial use.

Unfortunately, we don't know how. Proteins are linear chains of amino acids (of which there are 20 natural varieties), and we understand pretty well the correspondence between the genetic code—the DNA sequence—and the protein that's produced. But what's not clear to us is how that chain of amino acids folds up into a three-dimensional structure. To a first approximation, if a given amino-acid sequence folds at all, it will always fold to the same shape, even in different environments. The folding information is somehow encoded in the amino acids. We would love to be able to predict how a given sequence will fold, but we cannot with any degree of reliability. And even more relevant—and much more complicated—is the question of how that three-dimensional structure and specific array of amino acids determines what that enzyme does—what reaction it catalyzes and how well it does it. We can't tell whether one enzyme is better than another in any of its properties just by looking at it.

Besides being the scaffolding that turns a few critical amino acids into a very precisely shaped pocket that catalyzes a specific reaction, an enzyme's structure and sequence also determine its sensitivity to heat and cold, its interactions with other molecules that turn it on or off, its stability in various solvents, and all its other properties. Making subtle changes in the scaffolding quite some distance away from the catalytic pocket in an attempt to engineer one of these other properties can alter the enzyme's reactivity. Sometimes this confers the ability to perform the same reaction on a new molecule, or to catalyze a different reaction; usually it just makes the enzyme sick, that is, less active, less stable, or both. Despite decades of intense research into these protein-structure-function questions, we're not even close to having enough information to design any given enzyme "rationally."



Left: A "black smoker," a type of geothermal vent on the ocean floor. This one is about a mile and a half deep. Long assumed to be barren, black smokers harbor entire food chains whose metabolisms are based on such things as hydrogen sulfide, ammonia, and methane. Photo by John Barrows, University of Washington, courtesy of Diversa Corporation.

To complicate matters more, the molecule is so flexible that the X-ray crystallographic data on which we depend for structural information often can't tell us what's really going on. Furthermore, enzymes are constantly teetering on the brink of conformational disaster. A large number of forces—hydrogen bonding, electrostatic interactions, interactions with solvent molecules, and what have you—stabilize the catalytically active three-dimensional structure. But an almost equally large army of forces is working to unravel it, including competing interactions with the solvent and the entropy cost of folding it up in the first place. The net energy holding the molecule in the folded, active position is perhaps the equivalent of only two or three hydrogen bonds, compared with the hundreds of hydrogen bonds in the folded protein. (For people who like numbers, this is a few kilocalories per mole, or about 8–17 kilojoules per mole.) This is a real problem for the protein engineer, because when you start monkeying with the structure it's easy to make it unravel altogether.

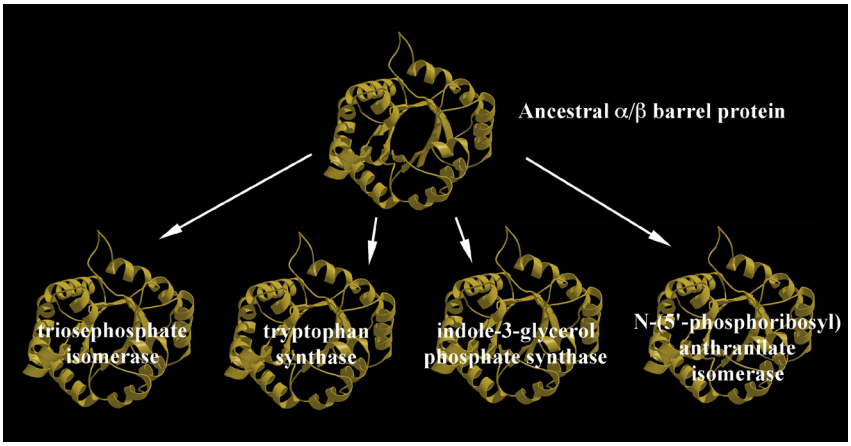
Now if the situation were really that grim, we could all just go home. Luckily, the protein-design problem is being solved all the time. As we speak, nature is creating proteins with novel functions in response to adaptive pressures. We can determine the amino-acid sequences of large numbers of modern proteins and, based on their degree of similarity, we can draw family trees that trace them back over hundreds of millions of years of divergent evolution. One ancestral protein can give rise to a huge family of proteins that look about the same but do many different things. Random mutation, recombination, and natural selection—evolution's "blind watchmaker"—have discovered fresh amino-acid sequences that confer new functions while conserving the overall three-dimensional structure.

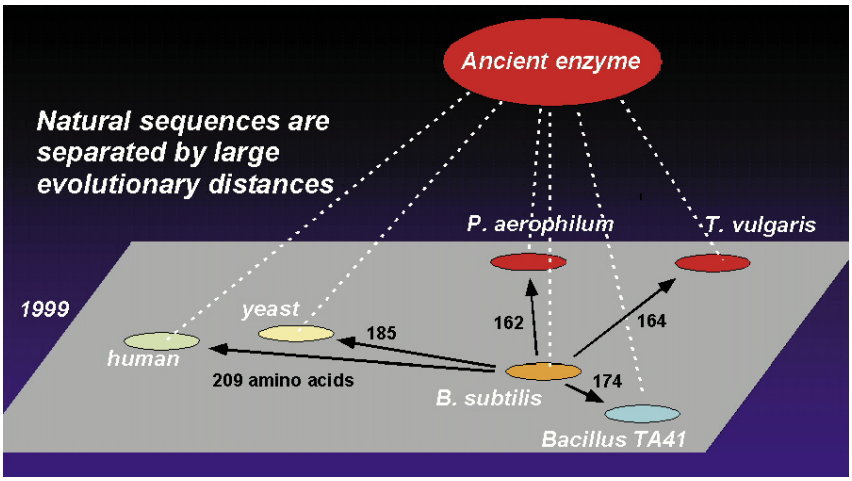
Molecules can evolve a lot faster than you might appreciate. A phosphotriesterase has recently been found that degrades at diffusion-limited rates—as fast as a catalyst possibly can—pesticides and biological-warfare agents that were invented less than 50 years ago. It's quite possible that this enzyme has come into being in just the past few decades. On an even shorter time scale, as those of us who have children know all too well, all sorts of illness-causing bacteria are evolving drug resistance in response to the large amounts of antibiotics we throw at them. This is survival of the fittest at the molecular level—drug-defeating enzymatic changes allow the bacteria containing them to live long enough to reproduce. This is why it's so important to finish taking your medicine even if you feel better—you want to kill all the bacteria, not just the weak ones.

Molecular evolution also helps life occupy diverse environments. A volcanic feature called a solfatara is essentially boiling sulfuric acid—pH 0 and 95° C—yet it's home sweet home to the microorganisms that are just teeming there. And there's life under the sea ice around Antarctica at −1.7° C. Molecular evolution has given rise to enzymes that are perfectly happy under these extreme conditions. This makes engineers like me really envious, because these are some of the attributes we'd like our industrial enzymes to have. (For a closer look at some unlikely places where life thrives, see the article by Ken Nealson on page 30.)

You might think that comparing the sequence of, say, a heat-loving enzyme and a lower-temperature one that performs the same function would help us figure out what mutations to make, but life's not that simple. Above right is a slice through the bacterial family tree showing various relatives of subtilisin E, a protein-cleaving enzyme that works at body temperature. The corresponding enzyme from *T. vulgaris*, a bacterium that lives in volcanic vents, is different by 164 amino acids—59 percent of its sequence—and we have no idea which of those substitutions are really responsible for its high heat stability. Most substitutions are neutral mutations that neither help nor harm the organism. This is called genetic

Below: Molecular evolution in action. Of all the enzymes whose structures are known, approximately 10 percent belong to the family of α/β barrel proteins, and presumably all evolved from a common ancestor. Although the progeny look alike to the casual observer, they do very different things.





Above: Subtilisin E, produced by *B. subtilis*, and its cousins produced by some other organisms. All of these enzymes break down proteins—in fact, subtilisin is widely used as a stain remover in laundry detergents. The numbers along the arrows indicate how many amino acids in each enzyme are different from subtilisin E.

drift, and it goes on all the time. Natural evolution is filled with historical accidents—random genetic drift—on which a little bit of adaptive evolution is superimposed.

We have discovered that the way out of the enzyme engineer's predicament is to look to nature—not for the specific molecules she has already made, but for the *process* she uses. Our challenge is to recreate, and direct, the evolution of molecules on time scales of less than hundreds of millions of years, because experiments of such duration really distress the grad students and are quite difficult to get funding for. For us, the maximum time unit for evolution is the PhD thesis—four years. But we'd really like to evolve new molecules in months or even weeks, which is now becoming possible.

Evolution may sound easy—just make mutations and see what happens. But that's not the case if you care about where you're going. Without a good strategy, your experiments are doomed to failure. That's because a typical protein has some 300 amino acids in its chain, and, with 20 letters in the amino-acid alphabet, there are 20^{300} ways to string those letters together. That's huge beyond imagination; huge beyond the number of protons in the universe. And this sequence space, if you will, is mostly empty—at least, mostly empty of the function you're interested in. So if you just wander around willy-nilly, it's not going to be a very useful exercise. For that reason, we do what nature does—we carry out local explorations of the space around existing, functioning molecules. We jump a natural enzyme through a new hoop, and accumulate mutations that help it jump higher.

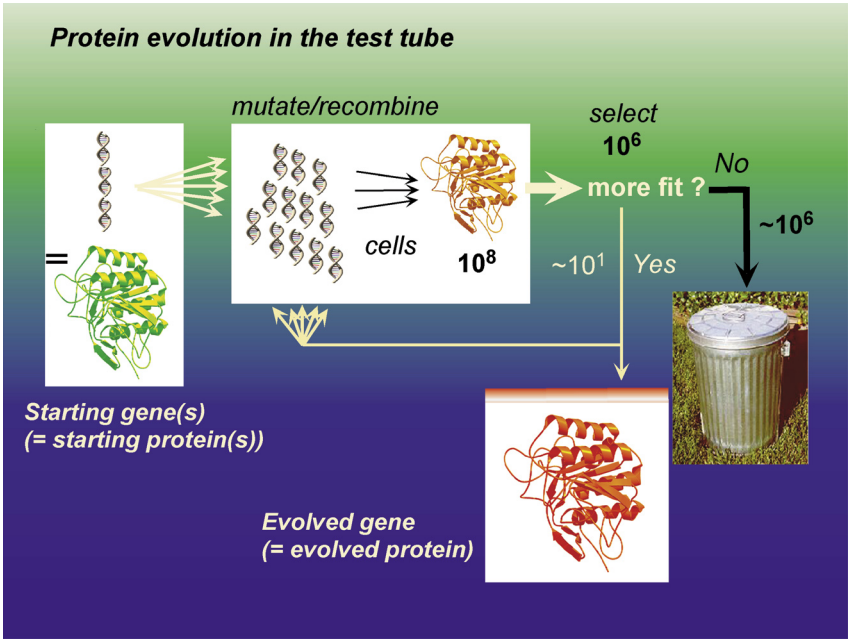
Just how local should this exploration be? If you plot, in the space of all its possible sequences, an enzyme's ability to perform a certain function (ignoring the fact that there are far too many dimensions to do this literally), the natural enzymes would be fog-shrouded mountain peaks, with the ground sloping away from them in (almost) all directions. If you take baby steps into the fog, however, you might discover that the peak is really a shoulder, and that the ground shortly begins to rise in one direction. But if you take a running leap, you're most likely to fall into a bottomless chasm. We find that the paths you discover taking small steps can often take you higher, sometimes much higher.

We have three major considerations in developing this experimental strategy. One is the first law of directed evolution: you get what you screen for. In other words, your success depends on how well your screen measures what you really want. The properties that you can measure easily may not be all of the ones that are important for the use you have in mind, so you either have to figure out some way of deducing the properties you're interested in from the properties you can measure, or you have to invest the time it takes to develop a new measurement(s). If you scrimp on this step, you'll wind up getting stuff you don't want. If you ask an evolutionary system to solve a problem for which trivial solutions exist, you'll get them, because they're easier to find. And if you just screen for one property (for example, the one you can easily measure), you may get a useless enzyme because it won't do the other things it's supposed to do.

The second consideration is that ways to improve a given enzyme are few, because enzymes are already so finely tuned. You could call this Murphy's law of evolution: most paths lead downhill. Even if you're asking the enzyme to do something completely new, most of the ways you can mutate it will make it worse. Beneficial mutations are rare, and combinations of beneficial mutations are extremely rare. So in order to find them, you have to do a pretty exhaustive search of your chosen area.

The third is that we have to screen enzymes by individually testing each mutant, one by one, to see how well it performs the combination of jobs that we're looking for. Given current technology, if we're clever and have a well-designed screen, we can maybe look at a million varieties of a particular enzyme per generation. This may seem like

Random mutation, recombination, and natural selection—evolution's "blind watchmaker"—have discovered fresh amino-acid sequences that confer new functions while conserving the overall three-dimensional structure.



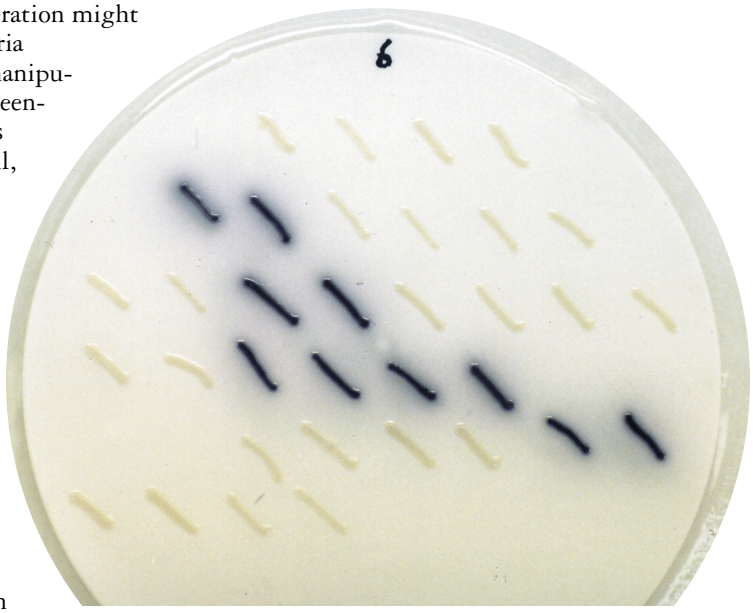
Above: Protein evolution by the numbers. Right: Colonies of bacteria on a petri dish. All the bacteria in any one colony are genetically identical to each other. The ones with the enzyme that performs the reaction we want have changed color.

a large number, but it's very small compared to sequence space.

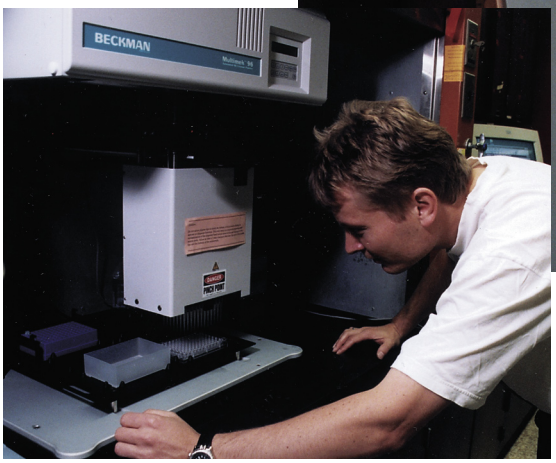
Together these considerations force us into the conservative, baby-step strategy—randomly changing only one or two amino acids at a time. There are 5,700 ways to change just one amino acid in a 300-amino-acid protein, 16 million ways to change two amino acids, and more than 30 billion to change three. The numbers grow so fast that it's impossible to search a reasonable fraction of even three-mutation sequence space, so instead we take a random walk of one- or two-mutation steps. This sounds slow, and pretty uninteresting, but what makes it all worthwhile is that mutations can be accumulated, either over many generations or, as I'll explain shortly, by recombination—a test-tube version of sex. And slow may not even be so bad, because one generation might take only a week or two. (The bacteria multiply overnight, but the DNA manipulation takes a day or two, and the screening takes the rest of the time. That's usually the bottleneck.) In a nutshell, we work with mutants that are very similar to their parents, and to do this we have to have a screening method that can measure small improvements in different functions simultaneously. Then we have to be able to accumulate these changes in order to make interesting, new enzymes.

So how does the experiment actually work? First, we isolate the gene that codes for the enzyme of interest. Then we mutate that sequence of DNA in a test tube, using the polymerase chain reaction. PCR, as it's called, can copy a piece of DNA very, very fast. While we are doing this, we introduce mutations at a specific rate by forcing the copying catalyst (an enzyme!) to make mistakes. We get it a little bit drunk, if you will, by adding metal ions to the mix. So we get a bunch of sloppy copies and create a library, so to speak, of mutants. We insert each of these mutated genes back into a circular piece of double-stranded DNA, called a plasmid, which has all the information that a bacterium needs to translate the DNA into protein. Each plasmid with its different mutation(s) goes into one bacterium, so now we have several million bacteria, most with a slightly different gene than the one we started with. We pour them out on a petri dish (the bacteria are suspended in water and spread out on the nutrient-rich surface, so that each bacterium is physically separated from the others by dilution), where they grow and divide until you can actually see, with your naked eye, individual colonies of genetically identical bacteria—in other words, colonies of clones. You use a robot, or you hire a bunch of undergraduates who sit there with toothpicks, to transfer each colony into its own well on the assay plate. Even better, you measure the enzyme activity right there on the plate by adding a reagent that changes color or fluoresces when the reaction occurs and then taking a picture. Then you screen to find the mutant that's most improved, extract its DNA, and start the process all over again. You stop when you have the desired enzyme (or when it's time for the student to graduate).

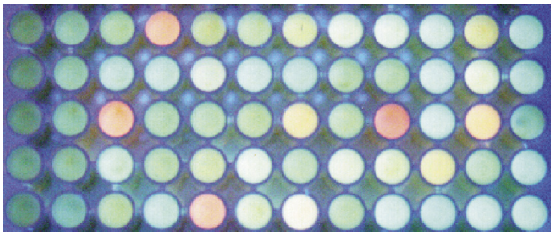
What I've just described is evolution by random point mutation, but there are other ways, too. We like to spice things up by adding a little sex, for example. There must be some evolutionary advantage to sex, to make up for its obvious



There must be some evolutionary advantage to sex, to make up for its obvious disadvantages. How else can you account for half the population not contributing to bearing the next generation?



Top right: Grad student Lianhong Sun uses an eight-channel pipetter to add reagents to a 96-well assay plate. Right: Or you can do the same job faster with a robot, as postdoc Oliver May demonstrates. Far right: Either way, you hope to get a reaction. Here, different variants of one enzyme are making different products from the same starting material.

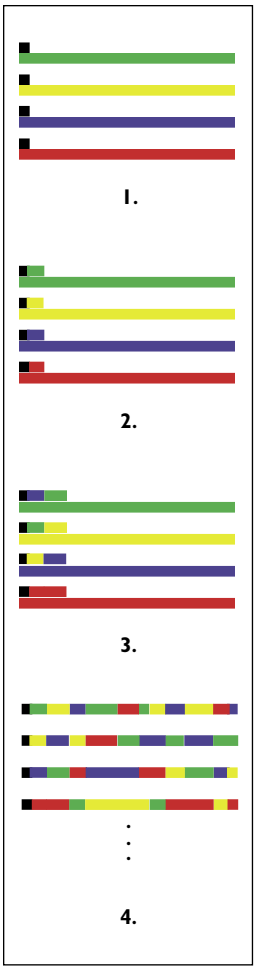


Left: The author's own products of evolution by recombination demonstrate their folding ability. From left to right are Joe (then one year old), James (eight), and Willy (two).

disadvantages. How else can you account for half the population not contributing to bearing the next generation? There are even things, like the peacock's tail, that are potentially harmful to an individual's survival. The compensation is that sex allows you to accumulate beneficial mutations from two parents at once, while flushing out the bad mutations. Molecular sex can be with any number of parents—sex with 50 even, if you can get all the genes to talk to each other. Sex is recombination, chopping up the genes and putting them back together in all possible combinations, so that now we're exploring a much larger (but still quite limited) region of sequence space. We can get the long legs and the thick hair from different parents, keep those good traits for the next generation, and throw out the undesirable offspring. Of course, it's much easier to do this with molecules.

How do you have sex in a test tube? Pim Stemmer, now at Maxygen, the company he

Below: Test-tube sex made simple. 1. Add a primer (black boxes) to several different versions of a gene (colored bars). 2. PCR begins copying the gene, starting at the primer. 3. When the primers come off and reattach, odds are they'll be on different genes than the ones they started on. 4. The result is a library of randomly shuffled genes.



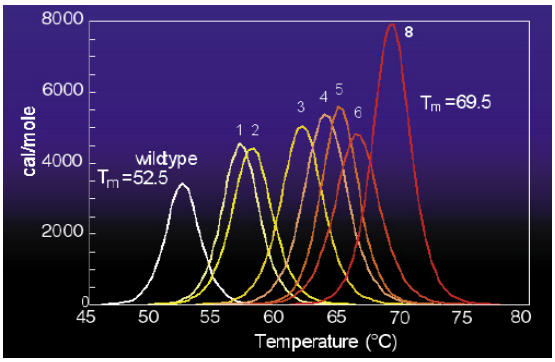
founded in Redwood City, invented this nifty method of gene shuffling that Huimin Zhao (PhD '98), postdocs Zhixin Shao and Lori Giver, my collaborator Joseph Affholter (who was at Dow Chemical but has since joined Maxygen), and I have improved. It's shown schematically at left. We put all the parental genes in a test tube and add a so-called primer, which is a short piece of DNA that initiates the PCR. The primer binds to a gene, and the PCR adds to the primer to make a copy. Normally, you'd let things run their course and get many complete copies. Instead, we heat the test tube after there's been time to process only, say, 20 letters or so, causing the primer and its unfinished copy to fall off the gene. When we cool the test tube back down, the primer latches on to the next parent gene it finds, and the PCR picks up where it left off. So if the primer was on the green gene initially, and landed on the yellow one in the second cycle, the copy will start with green information and continue with yellow information. And who knows—the next cycle might be blue. All this takes just a few minutes, and we end up with a library of what we call chimeric genes that contain randomly combined genetic information from the parents. (A chimera, in Greek mythology, had a lion's head, a goat's body, and a serpent's tail.)

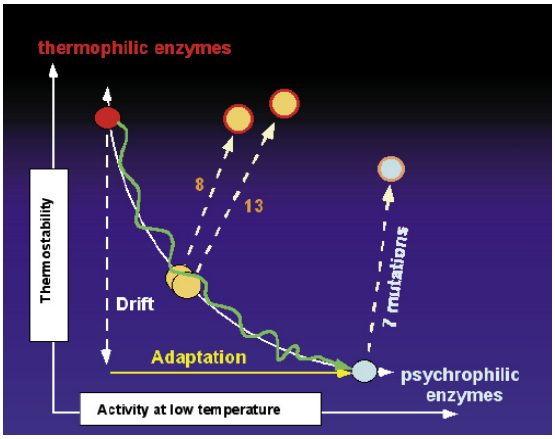
The next step is to find those rare good mutations and recombinations in your library. How you do this obviously depends on what you're looking for. I'll use an example of a thermophilic, or heat-loving, enzyme that we evolved. The enzyme is called para-nitrobenzyl esterase, because it breaks down the ester linkage in a family of compounds useful to synthetic organic chemists. We used a very simple screen on a 96-well plate (some plates have hundreds or even thousands of wells) to measure activity and thermostability at the same time. We made two copies of each master plate, and measured the initial activity on one copy after adding an ester that changes color when the enzyme cleaves it and seeing how fast the reaction proceeded. We roasted the second copy at a temperature high enough to cause the original enzyme to unfold, and then did the activity test on that plate. The mutants that passed became the parents of the next generation. Through a method called differential scanning calorimetry, in which we gradually heated the enzyme, we tracked how high we'd pushed the unfolding temperature. When a protein unfolds, it suddenly releases heat, so we measured the heat spike and noted the temperature.

Postdocs Lori Giver, Anne Gershenson, and Per Ola Freskgard did five generations of asexual point mutations and ended up with five parents that were the starting point for some test-tube sex, followed by a couple more generations of point mutations. The final result was an enzyme that didn't unfold until the temperature hit 69.5° C (an improvement of more than 17° and fully equivalent to a naturally "thermophilic" enzyme). But remember, you get what you screen for. Stability is relatively easy to improve, but it almost always comes at a price—usually in catalytic activity. Our screen allowed us to look for evolving enzymes that at least retained the low-temperature activity of the original one. But the activity of the evolved enzyme increases as you raise the temperature, so that it is actually more than 10 times as active at high temperatures than the original enzyme is at its preferred (lower) temperature. Consequently, we evolved better activity hand in hand with thermal stability. This is the sort of thing that gets industry really excited.

Natural enzymes are usually optimized for the temperatures at which their organisms grow—the heat-loving enzymes don't work well at room temperature, and the room-temperature enzymes aren't stable when you heat them. So people have assumed that thermal stability and low-temperature activity are incompatible. They've even devised theories to explain it: a high-temperature

Right: Differential scanning calorimetry results for the evolved para-nitrobenzyl esterase. The original enzyme (white curve) unfolded at 52.5° C, but after eight generations of evolution in the lab, the unfolding temperature had climbed to 69.5° C.



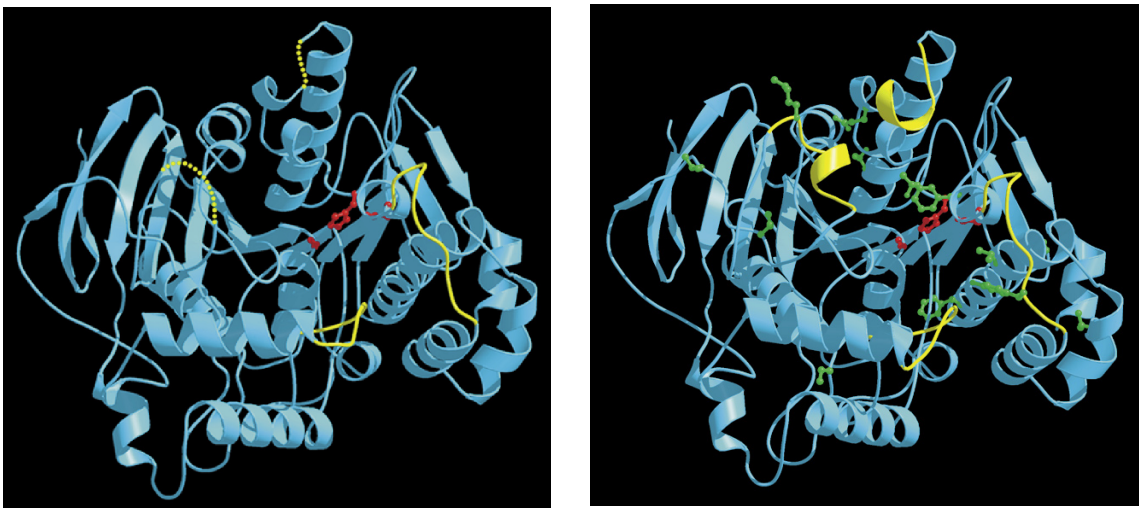


Top, left: Assuming that life began at high temperatures, a psychrophilic (cold-loving) enzyme could have evolved from a thermophilic (heat-loving) one as shown by the green arrow. There would have been no incentive to preserve thermal stability, so it would have slowly drifted away. But with the proper choice of selection pressures, or screens, enzymes that are both thermostable and very active at low temperatures can be evolved with surprisingly few mutations (numbered arrows). Left: Unlike the crud growing in most dorm fridges, this stuff is all being saved on purpose. Postdoc Anna Marie Aguinaldo takes stock.

enzyme has to be more rigid, so it will hold its shape, but a low-temperature enzyme has to be more flexible in order to guide the reactants when there's less energy available to the system. But here's another possible explanation: nature doesn't give a hoot about this combination of traits. Heat-dwelling organisms don't need activity at low temperature, and Antarctic bacteria certainly don't need heat resistance, so why go to the bother of making enzymes with both? If life began at high temperatures, as many scientists now believe, thermal stability may have been lost as enzymes more active at low temperatures evolved—because thermal stability wasn't needed any more, it drifted away. In the laboratory we can shed these biological constraints and really explore the difference between what's biologically possible and what's physically possible. We've found, to our delight, that a number of different properties are evolvable independently, which allows us to make very useful enzymes.

In the course of all this engineering, we can also study how a function evolves. During an experiment, we save all the intermediate mutants in the fridge. Once we've been successful, we go back and sequence the genes and identify the mutations that gave rise to the desired function. Here we found that only 13 amino-acid substitutions created this heat-loving para-nitrobenzyl esterase. We're not confounded by the hundreds of changes due to random genetic drift that would happen in a naturally evolving protein. In the laboratory, almost all evolution is adaptive, so we know that those 13 amino-acid substitutions are really responsible for the changing function. Knowing this, we can try to coax out the molecular mechanisms by which that property came about.

This is easier said than done. In fact, evolving a new enzyme is much easier than trying to figure out how it happened. Sometimes we ask for professional help. Professor Ray Stevens and his graduate student Ben Spiller at UC Berkeley



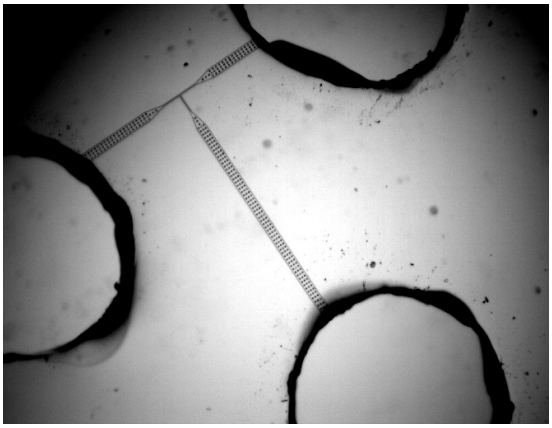
The evolved para-nitrobenzyl esterase (far right) and the original enzyme (right) look very much alike. The dotted lines are informed guesses—those portions of the enzyme were invisible to X-ray crystallography, presumably because their shape kept changing.

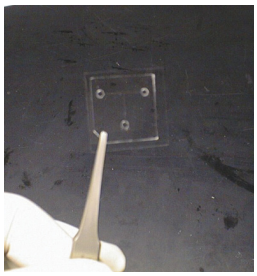
determined the three-dimensional structures of our evolved para-nitrobenzyl esterase and its progenitor. The structures are shown above. The red amino acids are the catalytic ones in the pocket where the ester binds. The sites of the 13 amino-acid substitutions are shown in green. It's fascinating to see how this enzyme has adapted. At first glance it might seem that nothing much has happened—the evolved enzyme folds up in pretty much the same way as its less stable and less active ancestor. But closer inspection reveals a number of interesting changes. What were two floppy loops (the dotted yellow lines in the ancestor's structure) have become fixed in the evolved enzyme. Mutations outside these loops in an early generation caused them to become rigid and added 11 new hydrogen bonds. This region then became a platform for further mutation later on. Two other loops (the solid yellow lines in the ancestor's structure) that control access to the catalytic site have also changed structure in the evolved enzyme. Note that most of these mutations are some distance away from the catalytic site. It would have been extremely difficult to predict them in advance. While we can rationalize the effects of each mutation after the fact, unfortunately there are no rules or patterns of substitutions that we could use in a future rational-design process. Kind of like "Buy low, sell high," the rules we generate are obviously true, but difficult to implement.

You can tune virtually any property of an enzyme to make it more useful for biotechnology, or to try to understand how the enzyme works. Laboratories around the world now are doing this with enzymes used in everything from laundry detergents to cleaning up chemical pollution. But if you want to create something really different, maybe even something totally new, nature doesn't offer much guidance as to how to go about it. Making an enzyme do something completely new is kind of like the species problem: it's easy to see how incremental changes create new breeds, but how do you make a whole new species when you can't imagine what a common ancestor—a transitional form that would get you from one species to the other—would look like? The problem is that it will probably take numerous amino-acid changes, and many at the same time, to convert one enzyme into another. But once again, sex offers a possible solution. The usual definition of a species is whether it can only have sex with its own kind, but we don't have such narrow-minded limitations in the laboratory. Molecules can have sex with anybody they want—sex with monkeys and worms and slime molds, if they feel like it. There only has to be enough similarity in the DNA for the gene-shuffling reactions to work. It might sound funny, but there's actually good reason to combine genes from widely divergent species.

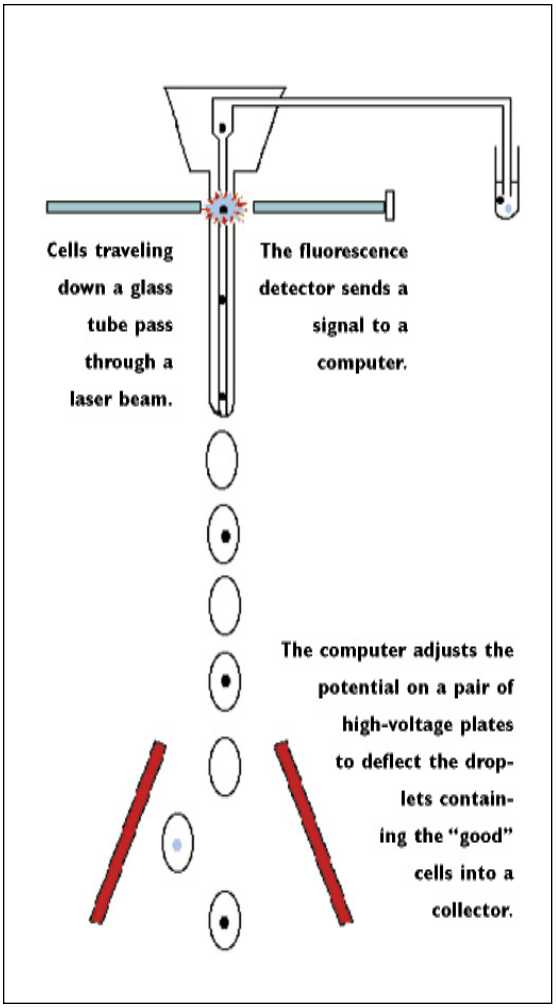
We have an enormous capacity now, through genome-sequencing studies and improved methods of gathering DNA in the wild, to identify genes for homologous proteins that have a variety of detailed differences in their sequences and functions. (Homologous means that the proteins come from a common ancestor and have essentially the same shape, even though their properties, or even functions, may have changed.) Pim Stemmer and his group at Maxygen showed that we can take homologous genes from various species and shuffle those genes to create a library of chimeras of enormous diversity, a fairly large fraction of which code for proteins that will still fold into the three-dimensional structure. (Remember, the hard part about making very large numbers of mutations is getting something that folds up properly in the first place.) But now that basic framework will be decorated with very different amino acids. These proteins can show a wide variation in properties, and possibly even completely new functions. (However, you do have to begin the process with some function in mind, in order to screen for

Right: A microphotograph of the T-shaped cell sorter. The black circles are the reservoirs for the junk cells, the keepers, and unsorted cells. The dots in the bottoms of the channels are pillars that hold up the roof.





Below: A conventional FACS uses electrostatic forces to sort cells (black blobs) as they fall from a glass dropper.

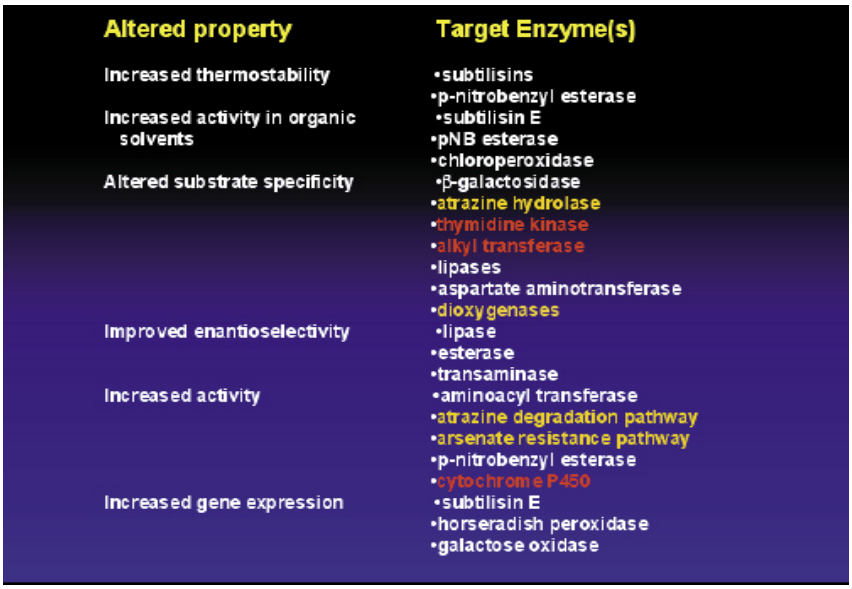


it.) So now, rather than exhaustively searching a little bit of sequence space close to the original enzyme, we're doing a sparse search in a vast but very special part of sequence space corresponding to folded proteins of the same overall structure. To return to the alpine analogy for a moment, we can leap from peak to peak like mountain goats. Of course, not all the peaks will be higher ones.

Moving on to even grander schemes, some day we'd like to be able to evolve whole new metabolic pathways—from a few to perhaps dozens of enzymes working in concert. To do so, we'll need to be able to look at enormous numbers of molecules, many more than we can look at today. We want to be able to evaluate 10^9 molecules, not just 10^6. Assistant Professor of Chemistry Rich Roberts has developed techniques he thinks will be able to look at 10^{13}. This is probably close to the upper limit for all practical purposes, just based on the mass of the molecules. It would get prohibitively expensive to work with much larger quantities of DNA.

In the meantime, our lab has been working with Associate Professor of Applied Physics Steve Quake's research group to develop a microdevice that can sort individual bacterial cells on the fly, based on their ability to carry out a reaction. The sorter is basically three wells and a covered, T-shaped channel, five microns wide and four microns deep (a bacterium is about one micron in diameter, and a micron is a millionth of a meter), cast in transparent silicone rubber from a silicon-wafer mold. The whole thing is about a centimeter square and fits on a microscope slide. The bacteria become fluorescent as a consequence of the reaction occurring. They enter at the base of the T, and as they flow, one by one, through a microscope's field of view, a computer reads the fluorescence signal and sends the cell into the T's left arm (the waste channel) or right arm (the collection channel) by changing the voltages at the channels' ends. This potential difference controls how the ions in the solution migrate, and the bacterium gets swept along in the current. Grad students Anne Fu and Charles Spence have built a prototype, shown above.

Conventional FACS (Fluorescence-Activated Cell Sorters) cost around $150,000, and they're terrible for working with bacteria. For one thing, they are very easily contaminated and take hours to clean out. Our plastic devices would cost a few cents and be disposable—you'd buy them in sterile pouches, like Band-Aids, use them once, and throw them away. Conventional FACS are also usually built to sort much larger cells from higher organisms like yeasts, plants, and animals, and have a hard time seeing bacteria. Our system will also be able to see the fluorescent bacteria much more easily. And conventional FACS imprison each cell in its own water droplet, which falls through a set of deflectors, so that you only get one pass through the sorter, while our microdevice is completely enclosed and lies flat. Steve's lab plans to exploit this by developing sorting



Above: Some of the enzymes that have been altered by directed evolution to date. The enzymes shown in yellow are designed for cleaning up various kinds of pollution problems, while the ones in red are intended to suppress unpleasant side effects of cancer therapy.

strategies that run faster than the switching speed. For example, we can run the machine really fast while shunting everything into the waste channel, and then when the rare good mutation zips by, we can quickly shift into reverse and draw it back out into the collection channel. It's like fast-forwarding through the commercials on your VCR and overshooting the point when the show comes back on—you just rewind a bit and catch it on the second pass.

The future of laboratory evolution is very bright, even though the field is only a few years old. Evolution is becoming a hot ticket because it works really, really well to solve problems that people care about. In fact, lots of labs in industry and academia are doing it. In a way, directing evolution is really a very old idea—animal and plant breeders have been doing it for thousands of years, albeit at much slower rates. Even with plants, you can generally only raise two generations a year, and you can only use two parents at a time. It's much nicer working with an organism whose population doubles every 20 minutes, and a gene that can have any number of parents. But it's a different way of thinking for many engineers and scientists, who aren't used to doing millions of experiments in hopes of finding one that works. Our knowledge is puny compared to what would be required to design enzymes from first principles, but if we settle back and admit our ignorance, it really frees us up to take this very different approach. These molecules are going to evade our understanding for quite a while yet. But when we use evolution, the lovely thing is that out come molecular solutions that *are* outside our understanding. So the future is no longer limited by our ignorance, it's really only limited by our imaginations. □

As *E&S* went to press, a paper written by postdocs Hyun Joo and Zhanglin Lin and me appeared in the June 17 issue of *Nature*. The paper describes the evolution of a cytochrome P450 that is much simpler than the natural enzyme. Cytochrome P450 is of interest to chemists because it inserts oxygen atoms into a huge number of compounds, but it's complex and ill-behaved. It needs a retinue of helper proteins and molecules called cofactors in order to work, and these guys are either impossible or very expensive to reproduce outside a cell. However, our P450 doesn't need any such help.

It's been known that hydrogen peroxide allows P450 to work unaided, so we turned this biochemical oddity into the enzyme's primary reaction pathway. Our version only took two generations to evolve and is 20 times better than the original one, which came from *Pseudomonas putida*, a soil-dwelling bacterium that uses it as a "digestive aid" to eat camphor.

Ivan Claeys (MS '88, PhD '91) gets his start cleaning up in biotech.

Frances Arnold's career has traversed engineering and science in a seemingly random walk. She earned a BS in mechanical and aerospace engineering from Princeton in 1979, and a PhD in chemical engineering from UC Berkeley in 1985. She was a postdoc in chemistry at Berkeley and Caltech before joining the Caltech faculty in 1987, where she is now a professor of chemical engineering and biochemistry. Along the way, she has received the Office of Naval Research's Young Investigator Award, the National Science Foundation's Presidential Young Investigator Award, and a David and Lucile Packard Fellowship. Her research interests include designing strategies for in vitro evolution, developing high-throughput screening technologies for catalysts, and evolving interesting new enzymes (http://www.che. caltech.edu/groups/fha). *She particularly enjoys raising her own three products of evolution by recombination.*

This article is a chimera created primarily from a SURF (Summer Undergraduate Research Fellowship) seminar given last summer and a recent Watson lecture.