Breeding Better Enzymes

Courtesy of Frances Arnold

Frances Arnold received the Nobel Prize in Chemistry in 2018 for pioneering directed evolution methods to make enzymes for applications in sustainable chemistry across medicine, consumer products, agriculture, fuels, and chemical production. She is the Linus Pauling Professor of Chemical Engineering, Bioengineering, and Biochemistry at the California Institute of Technology. In 2021, President Joe Biden appointed her cochair of the President's Council of Advisors for Science and Technology (PCAST). Arnold also received the National Medal of Technology and Innovation from President Barack Obama. She spoke at the 2023 International Forum on Research Excellence on evolutionary processes to engineer enzymes to take on novel functions. Here she discusses her latest research with American Scientist *editor-in-chief Fenella Saunders. (This interview has been edited for length and clarity.)*

How do you define directed evolution?
When I started as a protein engineer back in the 1980s, no one knew the rules for obtaining a better protein—*better* meaning that it performs a function that a human being is interested in rather than what the organism that carries it is interested in. No one knew the mapping from amino acid sequence to function. And we still don't know that for enzymes. We don't know what code gives rise to a particular function. We now know how to read DNA; we can sequence it. We know how to write DNA; we can synthesize it. We know how to edit DNA. But we don't know how to compose DNA. But that is a problem that evolution has solved really, really well. So to circumvent our near-complete ignorance of how sequences and codes function in enzymes, I decided to use nature's process that has proven very useful: evolution. I direct the evolution of enzymes.

Directed evolution was enabled in the 1980s by the methods of molecular biology. I cobbled together the ability to make mutations in a gene with error-prone PCR [polymerase chain reaction, a method to make copies of DNA], the ability to incorporate those genes into microbes and express the mutant proteins using recombinant DNA techniques, and the ability to do semi-high-throughput screening. The first spectrometers that could look at a 96-well plate were coming out in the late 1980s. So I pulled together all those pieces and started screening libraries of mutant proteins, finding beneficial mutations, and then feeding those back into an iterative process that was an evolutionary optimization of the enzyme. And you know what? It worked really well.

Natural selection doesn't always produce the best adaptation, but rather ones that are good enough to survive selection pressures. Is that part of the reason why there's room for directed evolution to create new functions?
Well, evolution doesn't work if there's no room for improvement, right? Nothing works if you're at some optimum, then it's hard to go uphill. But then you change the criteria for fitness, which is what we do when we evolve an enzyme to do some nonnatural task for which the enzyme is generally less enthusiastic. You're dropping the fitness relative to some other peak in the landscape and creating room for further optimization. That's critical. We often find that it's hard to improve enzymes for natural functions or natural environments. But it's really easy to improve them to do something different from what nature is asking them to do.

Do you combine the broader knowledge you have of proteins and enzymes to target areas for mutation, rather than randomly introducing mutations?
People often are tempted to do that because they think they know the rules, or some piece of the rules. But that leaves out the surprises, right? So yes, we do often target the active site of an enzyme when we're trying to change substrate specificity. But we also find that when we make random mutations, we find additional beneficial mutations whose effects we see but cannot explain. You leave out the opportunity to learn and to be surprised if you don't look more widely.

Everybody's tempted to target mutations because it's painful to screen all those variants, and they think they know better than random. But sometimes we don't know better than random. That's the fun part.

When you find beneficial mutations you can't explain, does that indicate incomplete theory?
Let's put ourselves back into the 1980s. Theory was very limited, right? We had very limited tools to understand the effects of mutations. People would say, well, the surface of the protein is not very important for function, things like that. But they were more rules of thumb rather than any kind of detailed theory. I think that our rules of thumb were just plainly wrong. I capitalized on that by finding that beneficial mutations would lie in areas of the protein that people thought were not important. That cemented this idea in people's minds that evolution was a good teacher. It's still difficult to design good enzymes based on theory.

Is your finding similar to evolutionary computation in the sense that those programs sometimes develop pieces that don't seem vital, but the code doesn't work if they're removed?
Yes, I think we both share the wonderful surprise element. Both fields find solutions that are very hard to reverse engineer. In the protein field, the reverse engineering is called biochemistry: you try to understand why a sequence has a particular function and how that function evolved. That's just as hard to do for a laboratory-evolved enzyme as it is for a naturally-evolved enzyme, even though we have the adaptive mutations. We know what the mutations are, but trying to understand, much less predict, what they're doing is still challenging. Evolutionary computations have some of the same features.

How do you involve machine learning in your laboratory?

The great power of directed evolution is that it's robust, and it works pretty well for many problems. It's a very general procedure. The downside to directed evolution is that it's slow and painful. You have to make all these mutations. You then have to test all these mutations. You throw away most of the mutations because they're not beneficial, and you usually have to do a few rounds of this. It is tedious to perform.

The dream is that we would do a smaller set of experiments, possibly not with the same mutational patterns, take the data, then combine the data with what we know about proteins—multiple sequence alignments, evolutionary language models, structure, molecular dynamics simulations—and learn from it. We would use multimodal models to empower predicting what should be the next set of variant sequences we should make, based on what we learned in each round. That's the dream. That's what we're trying to implement. We've been chipping away at that problem for several years now, trying to test different machine learning processes to see which ones are advantageous and under which conditions. But we've also found that good, old-fashioned directed evolution is a great baseline and hard to beat for many problems.

There's been so much change in technology since you began this work in the 1980s, so have you had to adapt yourself as well as the processes based on the new tools that are available?

I love tools. I'm an engineer; I'll grab any new tool that's useful. There are a lot of tools that are not useful. So the real trick is to know which ones give you better information, make better libraries, etcetera, and I'll happily use those. For example, now we can do moderate throughput screening with mass spectroscopy, which we couldn't do in the past, and we can even use LCMS—liquid chromatography mass spectroscopy—to identify improved enzymes. We can make mutant enzyme libraries in different ways. Other people have developed very powerful tools of continuous evolution. They're not as general as the directed evolution I described in the 1980s or 1990s, but they're much more powerful for specific problems. So not only are tools being developed, but also whole new modalities using evolution for design.

With directed evolution, is there still a place for a bottom-up process that builds proteins by design?

For directed evolution you always need to have a starting point, and that starting point has to exhibit enough of the desired functions, or something close to the desired functions, that you can measure so you can improve it. We're often stymied in finding that starting point. So one of the possible big benefits of protein design is to create a really terrible enzyme that could be a starting point for improvement by directed evolution. So there's an example where the two technologies blend beautifully. Design also takes you out into sequence space that's really different from what we find in nature. It can also take you to different function space. The real challenge today is how to design an enzyme or protein that has the desired novel function. If you can do that, you could presumably use that designed enzyme for improvement by evolution.

"The more we demand environmental friendliness and sustainability for our chemical processes and therefore put a price on producing waste, the more we're going to see enzymes used in industry."

How can directed evolution help with the amount of waste produced in the chemical industry?

This is the main focus of my research. For me the glory of enzymes and the microbes that produce them is that they can use renewable abundant resources and convert them into a vast panoply of beautiful chemical products and materials. We're just scratching the surface of training microbes or evolving their enzymes to do that for us. There are a few great examples of chemical processes that have been completely replaced with enzyme-based processes. Merck, for example, has won U.S. EPA Green Chemistry Challenge Awards because they've demonstrated a dramatic reduction in waste production for making pharmaceuticals. They've gotten past the use of toxic metals by using enzymes. Microbes will produce these marvelous catalysts for sugar, and I just find that a wonderful vision for the future.

Do you see widespread usage of these enzymes happening anytime soon?

Well, enzymes are already a big business. A lot of enzymes show up in consumer products. There are all the enzymes in your laundry detergent, all sorts of enzymes in various manufacturing processes, enzymes and microbes in ethanol production, or in making high fructose corn syrup. All sorts of big-scale industrial processes use enzymes.

The more we demand environmental friendliness and sustainability for our chemical processes and therefore put a price on producing waste—there's a real cost of that waste to society, so there should be a cost on producing it for anybody who uses those products—and the more that we make producers internalize those costs, the more we're going to see enzymes used. People will just naturally transition to that.

That's why, for example, Merck is transitioning to biocatalysis, because they do not want to produce hundreds of kilograms of waste for every kilogram of drug product. The cost of treating that waste is high. So I see this change happening now. We're at the very beginning of that transition. But I'm hoping over the next 10 or 20 years we'll see a lot more processes currently done with bad chemistry switch over to good enzymes.

Can directed evolution make enzymes that function in extreme environments?

Yes. Enzymes that degrade PET [polyethylene terephthalate] plastics have been evolved to be much hardier, so they can work at higher temperatures or under more extreme conditions, in an industrial degradation process. Cellulases have been made more thermostable so that they can function to degrade cellulose into sugars under conditions in which the viscosity of the feedstream is lower and the kinetics are

higher. Almost all industrial examples and processes benefit from making the enzymes hardier, because then they also tend to last longer, which reduces cost. I mean, even the enzymes in your laundry detergent have to be hardy, right? Because who wants to work in your laundry machine with surfactants and bleach and all that? Those enzymes have to be hardy little suckers.

Carbon sequestration also is a really nice example. The enzyme company Codexis demonstrated that they could make a carbonic anhydrase that functioned at close to 100 degrees Celsius in which the process was much, much better, and it was pretty surprising how nasty an environment that enzyme worked in. Enzymes can be a lot hardier than people give them credit for. Of course, they can also be very weeny—they can be really terrible sometimes.

Are evolved enzymes being used in renewable energy production?

I cofounded a company called Gevo in 2004 to make liquid fuel from renewable plant resources. It went public in 2011, and is still in business. We had a technology to make isobutanol from corn and agricultural wastes. However, the price of oil dropped precipitously and the price of corn skyrocketed, so you can't make money on that. It's very hard to compete with pumping energy out of the ground when we're not internalizing the CO_2 cost. There's plenty of great technology out there. Renewable ethanol is now being used to make jet fuel, and there are many other possibilities. But the bottom line is that until we really incorporate the climate costs of making these fuels from oil, it'll be very hard for a biological process to compete.

But I think the future is incredibly bright for bringing biology and the chemistry of the biological world into our human chemical world, and vice versa! We can learn from what chemists have done, and bring that to teach biology how to do completely new chemistry. So we have a lot to learn from each other. It's a whole new field with much, much potential for societal good.

Your academic path was not the most linear, and you also seem to have no reservations about bucking convention. Do you think those tendencies helped you to be innovative in your research?

I think what helped me be innovative was my willingness to try all sorts of new things. I have an undergraduate degree in mechanical and aerospace engineering and a PhD in chemical engineering. I've had to learn some biology. I've lived all over the world. I took many different kinds of jobs before I went to graduate school. I wasn't afraid to learn new things, to do new things, and to look at problems in perhaps a different way from the standard. And honestly, if you're going to go out and do something totally new, you need to have some degree of fearlessness, and be able to take criticism and not crumble under the negative attention that doing something different often brings. So I think my doing things that were not the standard for women in the 1970s gave me an advantage later on, when I decided to do research in an area that didn't even exist.

A lot of people ask me about my path to a Nobel prize, as if that might inform their own path. I say that there is no one path—you do you. You have to be yourself and not like everybody else. I mean, for me, the fun path is not to be like everybody else. A good path is to embrace your differences and to be different. ■