

Chemistry Nobel winners harnessed evolution to teach old proteins new tricks

The combination of diversification and selection has proved a powerful recipe for creating enzymes and drugs.

For most of human history, the process of making things was purely empirical. Early bakers of bread knew nothing of the microbiology of yeast, the thermodynamics of ovens, or the Maillard reaction that creates a golden brown crust. The architects of buildings such as the Colosseum in Rome or the Hagia Sophia in Istanbul, Turkey, had no quantitative understanding of how forces propagated through their structures. But over generations of trying different things, they developed procedures that worked.

When structural engineers finally learned how to apply Newtonian mechanics, it launched a revolution in building design. Similarly, the advent of quantum mechanics and understanding of atomic structure enabled vast improvements in synthetic chemistry, and developments in chemistry brought new possibilities in food processing. Time and again, improved understanding of the physical world has opened up a fast track to making better things.

From that point of view, this year's Nobel Prize in Chemistry—half of which was awarded to Frances Arnold of Caltech, the other half jointly to George Smith of the University of Missouri and Greg Winter of the MRC Laboratory of Molecular Biology in Cambridge, UK—might come as a surprise. The Nobelists developed methods for making new proteins that largely bypass the need to understand the relationship between protein sequence, structure, and function. Not only do the methods work well, but they actually perform better—at least for now—than attempts to rationally design proteins for the same purpose. Their products include one of the best-selling pharmaceuticals in the world and enzymes to catalyze previously impossible reactions.

Arnold is the fifth woman, and the third not named Curie, to be awarded a chemistry Nobel. With Donna Strickland receiving a share of the physics prize (see the story on page 18 of this issue), 2018 is only the second time in history—the

first being 2009—that two or more women have garnered Nobels in any science category in the same year.

Eppur si evolve

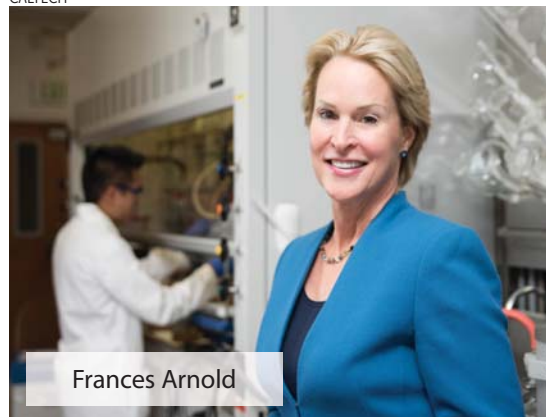
Although it's not yet possible to engineer most proteins the same way we engineer skyscrapers or computers—through confident rational design informed by physical understanding—it's not for lack of trying. Researchers across disciplines have been plugging away at the protein-folding problem, and they've developed a suite of computational and experimental techniques for analyzing protein structure and function. But gaps remain in their understanding of the details—and for functional proteins, especially enzymes that catalyze particular chemical reactions, the details matter.

Enzyme function is a delicate balancing act. The enzyme must bind and hold its target molecule in its active site, but it also needs to let go of the product molecule when the reaction is complete. That balance can be thrown off by tiny errors in the spatial positions of amino acids, far beyond the resolution of current methods for predicting protein structure. Even given an existing enzyme as a starting point for design, when all that needs to be done is swapping out a few amino acids to slightly change the catalyzed reaction, there's no way to reliably predict which substitutions need to be made to achieve the desired effect.

Fortunately, biology has its own way of making new proteins that's been tried and tested for billions of years. The proteins in the body of every living thing have been optimized for specific functions over countless generations of evolutionary trial and error, and new proteins are evolving all the time. Bacteria evolve resistance to antibiotics, for example, and microorganisms in polluted environments can even evolve the machinery to metabolize the pollutants as food.

It might seem strange that biology can do that. The space of all possible pro-

CALTECH



Frances Arnold

UNIVERSITY OF MISSOURI



George Smith

MRC LAB OF MOLECULAR BIOLOGY



Greg Winter

tein sequences is enormous; even for a relatively small protein composed of 200 amino acids, there are 20^{200} possibilities. (That's even more than astronomically many: The number of atoms in the known universe is a mere 10^{80} or so.) Moreover, the vast majority of those sequences constitute biochemical nonsense that doesn't encode any useful function. Searching that space for the best protein for a particular job is an example of a type of

math problem called combinatorial optimization. Other well-studied combinatorial optimization problems—including finding the energetic ground state of a spin glass or the shortest route among cities for a traveling salesman—are notoriously difficult to solve even with clever algorithms, let alone a blind search.

But protein evolution has a couple of points in its favor. First, it doesn't necessarily need to find the very best sequence—just one that's good enough for the purpose. Combinatorial optimization problems often have many near-optimal solutions that are configurationally very different from the global optimum. If the same holds true in protein space, any one of the near-optimal sequences could be a perfectly good evolutionary goal.

Second, although sequence space is large, it's also compact. Any protein of length 200 can be transformed into any other through at most 200 single-site mutations. Even a modest 10 rounds of mutation can explore a swath of sequence space large enough to contain proteins with completely new functions.

Even given all that, there's no *a priori* guarantee that evolution should work. It could be that functional proteins are so hidden from each other in sequence space that any path from one to another would pass through a completely non-functional intermediate—maybe one that doesn't even fold into a stable structure. On such a forbidding landscape, one functional protein could never evolve into another.

Proteins in nature, however, do evolve and change their function. It must therefore be the case, for whatever reason, that at least some functional proteins are clustered together in sequence space, with a fairly smooth local relationship between sequence and function. If nature can exploit that smoothness to build new proteins, researchers should be able to do the same.

Reactions come to life

In 1993, when most of the protein research community was still committed to the rational design approach, Arnold published her first account of directed enzyme evolution.¹ The principles are little changed since then. The process has two main ingredients: some way to induce diversification in a population of proteins, and some means of screening or selecting for the proteins in the group that are

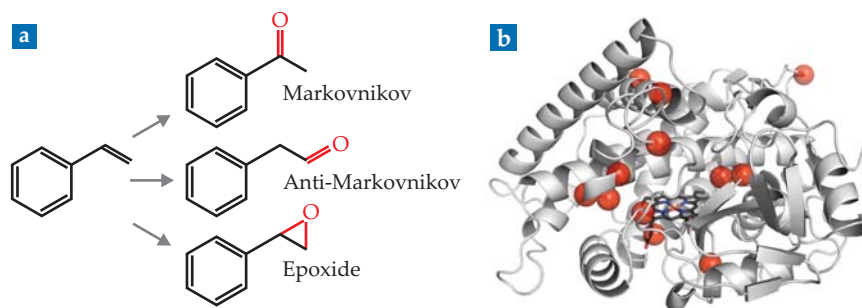


FIGURE 1. AN UNNATURAL REACTION. (a) Of the three ways an oxygen atom can add to a carbon-carbon double bond, the anti-Markovnikov reaction product is by far the hardest to produce. (b) An enzyme created by directed evolution catalyzes the anti-Markovnikov reaction as its primary product. The red spheres mark the 12 amino acids that were changed from the starting enzyme. (Adapted from ref. 3.)

closest to the desired function. Repeating those two steps, typically around 10 times each, often gives good results.

Several options exist for each step. The most straightforward way to create protein diversification is to take the DNA that encodes the starting protein and replicate it under conditions that purposely induce copying errors. If desired, the errors can be concentrated in a particular region of the protein, such as the area around the active site of an enzyme. Nowadays, with biotechnology capable of producing custom DNA sequences on demand, specific mutations can be induced at will. And in a sort of molecular equivalent of sexual reproduction, portions of the DNA from a population of "parent" proteins can be mixed and matched in different ways to create a diverse generation of "children."

Selection and screening strategies vary greatly in the size of the protein population they can handle. At the higher-throughput end of the spectrum are *in vivo* methods in which the function of the enzyme is somehow linked to the survival of the bacterium that produces it. Because the bacteria all compete with each other for survival without much effort on the experimenter's part, millions of enzyme variants can be screened in parallel. But the complexity of biochemical pathways means that the test can easily be fooled: Bacteria can evolve unintended ways of staying alive that are independent of the enzyme function.

Arnold often favors the lower-throughput method of placing each enzyme variant in a separate container and seeing what reaction products it makes. That's easiest if the products can be made to fluoresce or change color; then a whole array of enzymes can be evaluated at a

glance. But one can also manually evaluate the variants' reaction products one by one with time-tested analytical techniques such as chromatography or mass spectrometry. It's laborious to do that for 100 or more enzyme variants over each of 10 rounds of evolution, but when the outcome is a valuable new enzyme, it's a price worth paying.

Also important, of course, is a suitable starting point that's likely to have a smooth, not-too-long path to the evolutionary goal. "If you want to evolve a racehorse," says Arnold, "you don't want to start with a beetle." The first years of directed enzyme evolution focused on getting enzymes to perform reactions that naturally occur in biology, but under different conditions—a higher temperature or different chemical environment, say—that are more useful for industrial chemistry. In those cases, the researchers start with the enzyme that already performs the reaction.

More recently, Arnold and colleagues have been creating enzymes for reactions that are not present in biology. They do it by exploiting enzymes' so-called promiscuity, the ability to catalyze reactions other than the ones for which they're optimized. Because the enzymes don't naturally encounter the reactants for the nonbiological reactions, they never catalyze those reactions *in vivo*, and they don't usually catalyze them very well. But a few rounds of directed evolution can transform a poor catalyst into an excellent one.

Most of the nonbiological reactions the group has tackled—such as the formation of a carbon-silicon bond,² which despite the Earth-abundance of both elements appears nowhere in any living thing on Earth—are already common in

synthetic chemistry. But enzymes offer advantages over traditional chemical methods. Their building blocks are inexpensive and easy to come by, unlike costly precious-metal catalysts whose mining damages the environment. And like all biomolecules, enzymes exist in a world of molecular chirality, or mirror asymmetry, so they often produce one mirror-image form of their product molecules but not the other. That chiral selectivity is valuable in many contexts (see *PHYSICS TODAY*, July 2018, page 14), but it's hard to achieve with traditional synthetic methods.

Last year, Arnold's group even created an enzyme for a reaction that's perplexed synthetic chemists.³ Figure 1a shows the three ways an oxygen atom can react with a carbon-carbon double bond. The Markovnikov and epoxidation reactions are relatively easy to catalyze, but the anti-Markovnikov reaction, due to its unstable transition state, is disfavored. Arnold and colleagues found an epoxidation enzyme that also produced a small amount of the anti-Markovnikov product; with 10 rounds of directed evolution, they optimized it to produce the anti-Markovnikov product almost exclusively. The new enzyme is shown in figure 1b. The mutated sites, shown in red, are scattered across the molecule, with many quite far from the active site at the center. As is typical for their results, Arnold and colleagues don't know why those mutations have that effect. "It just works," says Arnold. "And we like things that work."

See what sticks

Directed evolution is not just for enzymes. Roger Tsien, in work that won him a share of the 2008 Chemistry Nobel, transformed natural green fluorescent protein into a rainbow of fluorescent proteins through repeated rounds of random mutagenesis. (See *PHYSICS TODAY*, December 2008, page 20.) Researchers today continue to use directed evolution to tune fluorescent proteins' photophysical properties.

The third major class of directed-evolution targets are proteins that inter-

act with other proteins. That category includes therapeutic antibodies, the subject of the other half of this year's Nobel. Independently of Arnold's enzyme work, Winter conceived a randomization procedure that exploited a selection technique developed by Smith and led to the discovery of the drug adalimumab, sold under the trade name Humira. For more than 15 years, Humira has been used to treat a range of autoimmune disorders, including psoriasis and rheumatoid arthritis. What makes the drug special is that it's the first therapeutic antibody whose source is fully human. More are in development.

Antibodies from animals have been used in human medicine for more than a century. Like small-molecule drugs, they mostly work by binding to proteins in the membranes of misbehaving cells—whether those cells shouldn't be there at all, like pathogens or cancer cells, or whether they're otherwise normal cells that are doing something harmful, like stimulating the immune system to attack the body's own tissues. Nonhuman animals such as mice, when injected with human cells, recognize those cells as for-

eign and send an army of diverse antibodies to get rid of them. If one of those antibodies binds to the membrane protein of interest, it can be replicated and used to treat patients.

A major advantage of antibody drugs is that they're large proteins with complex shapes. They're a lot pickier about where they bind than small-molecule drugs are, so they potentially have fewer side effects. The disadvantage of animal antibodies is that the human body recognizes them as foreign and purges them with its own immune system, often before they can be fully effective.

In the 1980s several groups developed strategies to reduce the immune response by making antibodies that are part human, part animal. So-called chimeric antibodies fuse the body of a human antibody to the target-binding lobes of an animal antibody; they include the autoimmune drugs Remicade and Rituxan. Humanized antibodies—Winter's approach—go a step further: They're almost entirely human, with just

the very tips of a nonhuman antibody attached.⁴ They include the cancer drugs Avastin and Herceptin.

Not long after he developed the humanization protocol, Winter wondered if he could do even better and create therapeutic antibodies that are wholly human. It would be highly unethical to inject human subjects with, say, cancer cells just to harvest and clone their antibodies. So Winter took what he calls a "master thief" approach: given a large enough collection of keys, or antibodies, there's a good chance that one of them will fit in a given lock, or target protein.

To make the keys, he mixed and matched portions of existing human antibodies. "The immune systems of vertebrates use a similar strategy to make antibodies to a wide range of antigens," explains Winter; Susumu Tonegawa was awarded the 1987 Nobel Prize in Physiology or Medicine for discovering that antibody gene segments are naturally rearranged into a nearly endless number of combinations. "We aimed to mimic that strategy using the same antibody building blocks." From an initial library of 1000 antibodies, 1 000 000 combinations

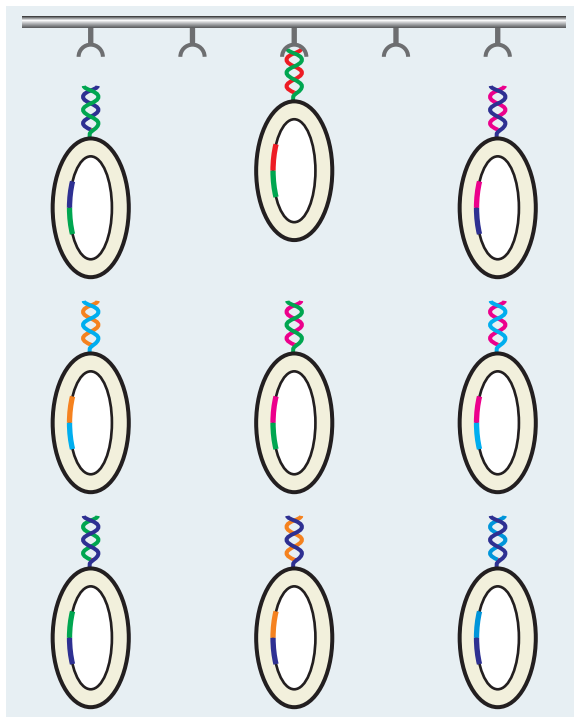


FIGURE 2. PHAGE DISPLAY FOR ANTIBODY SELECTION.

Bacteriophage viruses (black ovals) are endowed with random combinations of human-antibody gene fragments (interior colored segments) and display the resulting proteins (colored chains) on their surfaces. Target proteins affixed to a solid substrate are used to fish out the antibody that best fits the target.

can be made, of which 999,000 are new. That's far too many to test one by one. So how to determine which one, if any, fits in the lock?

That's where Smith's idea came in. In 1985 he showed that when a gene for a foreign protein is inserted into a bacteriophage, or virus, the virus not only synthesizes but incorporates the protein into its coat.⁵ The method, called phage display, is tailor-made for screening protein-protein interactions. A large population of phages, each displaying a different protein, can be washed over a solid substrate that's slathered with a target molecule. The phage whose protein binds to the target gets stuck to the surface, while all the others get washed away.

Winter used phage display as a massively parallel platform for screening his new antibodies. As sketched in figure 2, a population of phages was created with different random combinations of antibody gene fragments; whichever ones stuck to the target contained the right combination for building the new human antibody.

Humira is an antibody that targets tumor necrosis factor α (TNF α), a signal-

ing protein in the membranes of white blood cells that promotes inflammatory responses throughout the body. Inflammation can be part of normal immune system activity—the fever that accompanies the flu isn't caused by the flu virus itself but by the immune system trying to get rid of it. In patients with autoimmune disorders, however, TNF α needlessly causes debilitating inflammation in otherwise healthy tissues. Humira binds to the misbehaving TNF α and suppresses its activity.

Quo vadis?

Directed evolution has produced some impressive results, but is there anything it can't do? So far, the enzymes produced by directed evolution operate in aqueous solution, so they can't catalyze reactions that require a water-free environment. But that's a technical hurdle, not a fundamental limitation.

Even given the vast space of possibilities, there's no guarantee that a protein with a desired function exists—or if it does, that it's within striking distance of any accessible starting point. On the latter front, computational methods for de-

signing proteins from scratch may help. Although *de novo* design lacks the precision to create a finished, functional protein in most cases, it can often produce a reasonable starting point that can then be optimized by directed evolution.⁶ The combination of methods provides access to regions of sequence space that biology has yet to explore.

The outcome of a directed evolution process is only as good as the experimental procedure. One gets what one selects for, and an incorrectly conceived selection criterion can easily have unintended consequences. But with some persistence and the right experimental design, Arnold says, "you can get from here to some pretty cool places along a fairly smooth path."

Johanna Miller

References

1. K. Chen, F. H. Arnold, *Proc. Natl. Acad. Sci. USA* **90**, 5618 (1993).
2. S. B. J. Kan et al., *Science* **354**, 1048 (2016).
3. S. C. Hammer et al., *Science* **358**, 215 (2017).
4. L. Riechmann et al., *Nature* **332**, 323 (1988).
5. G. P. Smith, *Science* **228**, 1315 (1985).
6. J. Dou et al., *Nature* **561**, 485 (2018). PT

DUNIWAY

STOCKROOM CORP.

Supplying reliable
vacuum equipment
since 1976



Mechanical
Pumps



Ion Pumps



Turbo
Pumps



Gauge
Controls



Vacuum
Sensors



Hardware



Supplies



Diffusion
Pumps



www.duniway.com

800.446.8811 (Toll Free) 650.969.8811 (Local) 650.965.0764 (Fax)

PRECISION MEASUREMENT GRANTS

The National Institute of Standards and Technology (NIST) expects to make two new Precision Measurement Grants that start on 1 October 2019, contingent on the availability of funding. Further guidance will be provided on the Web when the funding level is resolved. The grants would be in the amount of \$50,000 each per year and may be renewed for two additional years for a total of \$150,000. They are awarded primarily to faculty members at U.S. universities or colleges for research in the field of fundamental measurement or the determination of fundamental physical constants.

Applications must reach NIST by **1 February 2019**. Details are on the Web at: physics.nist.gov/pmg.

For further information contact:

Dr. Peter J. Mohr, Manager
NIST Precision Measurement Grants Program
100 Bureau Drive, Stop 8420
Gaithersburg, MD 20899-8420
301-975-3217

NIST

National Institute of Standards and Technology
Technology Administration, U.S. Department of Commerce