CASE STUDY

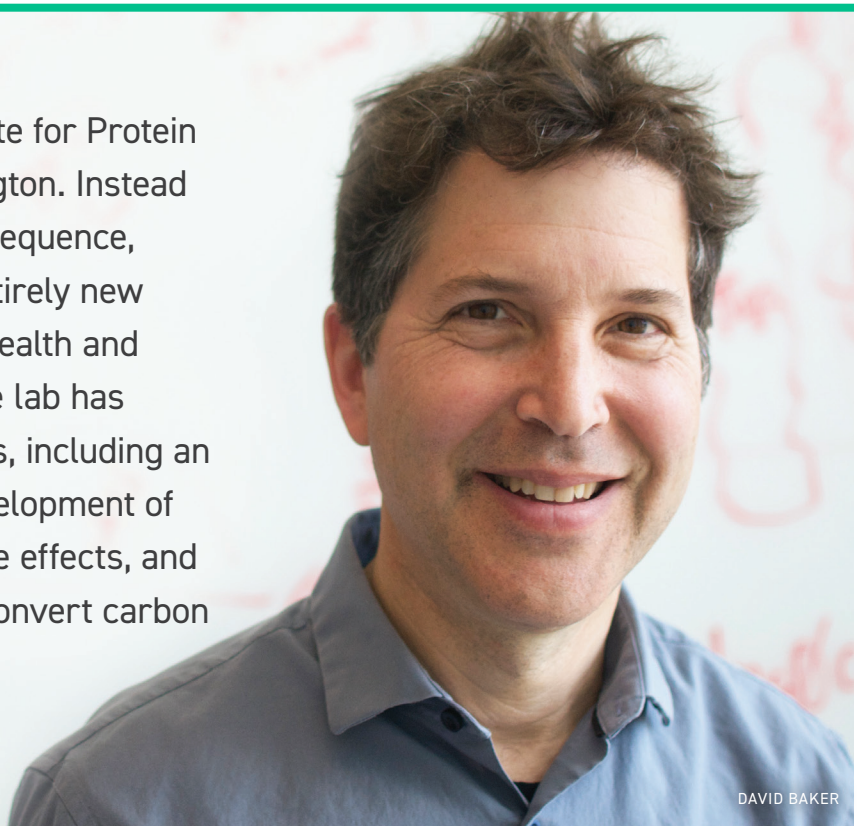
TWIST BIOSCIENCE

THE BAKER LAB AT THE UNIVERSITY OF WASHINGTON

Designing Proteins to Revolutionize Biotechnology

MAY 2019

The Baker Lab is part of the Institute for Protein Design at the University of Washington. Instead of engineering an existing protein sequence, the lab focuses on the design of entirely new proteins to address challenges in health and technology. Led by David Baker, the lab has registered a series of achievements, including an experimental RSV vaccine, the development of anti-cancer proteins with fewer side effects, and new enzymes that help microbes convert carbon dioxide into useful chemicals.



DAVID BAKER

Recently, Baker Lab received a $45 million grant from the Audacious Project, a funding consortium, to pursue several computational design projects, including a flu vaccine capable of providing lifetime immunization, advanced protein containers for targeted gene delivery, and smart proteins capable of identifying cancerous or otherwise unhealthy cells.

Despite these advances in protein design, progress has been difficult because the majority of designed protein sequences fail to fold or don't function as designed. The computer models used for protein design produce many "false positives" (sequences that look good on the computer but fail in the lab), and understanding why these failures happen is very challenging. A protein might fail for many hundreds of reasons, and a failed design can provide little feedback on what went wrong. Recently, the Baker Lab sought to overcome this challenge through a combination of computation and large-scale experimentation. Testing designs at a much larger scale makes it possible to look for patterns in both the successes and failures.

Evaluating Protein Folding with Oligo Pools

In two recent reports, Baker Lab researchers markedly advanced their protein design methodology using Twist Bioscience Oligo Pools. Oligo pools consist of between thousands and hundreds-of-thousands of synthetic oligonucleotide sequences, pooled together. In papers published in *Science* and *Nature*, the researchers used oligo pools to encode large numbers of potentially viable protein structures. Gabriel J. Rocklin, PhD, a former fellow at the Department of Biochemistry of the University of Washington and a member of Baker Lab, who was a lead author of both studies, said that by designing very small proteins that could be encoded by a DNA sequence the size of a single oligo, new protein designs could be tested in high throughput. Rocklin is currently an assistant professor of Pharmacology at Northwestern University in Chicago, and a member of Northwestern's Center for Synthetic Biology.

Achieving a Rate of Success Not Previously Possible

In one study, Rocklin and the team designed and screened thousands of proteins to determine whether each formed a stable, folded structure. Proteins that don't fold properly are extremely sensitive to proteolytic enzymes, enabling the team to use resistance to proteolysis as a measure of folding. The designed proteins were expressed in yeast, with each design being presented on the surface of the yeast cell that synthesized it. Cells expressing stable proteins were labeled with a fluorescent dye and collected in a cell sorter, so that DNA sequencing could reveal the molecular identities of all the stable designs.

"There are so many things that can go wrong with a protein," Rocklin said. "In the old design approach, you make 10 or 15 proteins, and three of them work while the others don't. That's great to get some success, but it doesn't allow us to improve in the future. Now, by testing 4,000 or 10,000 proteins at a time, we get enough data and experimental feedback to be able to improve our computational tools."

Encoding the larger number of proteins in affordable oligo pools gave Rocklin the information he needed.

"That was super useful feedback that went into improving the computational model and improving the design process," he said. "Large scale oligo library synthesis that's affordable enables

Our ability to get huge numbers of oligos affordably...made the experiments possible.

Gabriel J. Rocklin, PhD
ASSISTANT PROFESSOR, NORTHWESTERN DEPARTMENT OF PHARMACOLOGY AND CENTER FOR SYNTHETIC BIOLOGY

us to do something that I think protein designers had been wanting to do for a long time but had not had a way to do, which is test thousands of completely customized, computer designed protein sequences.

"The very first time we tested any of the designs, the overall success rate was about five percent," Rocklin said. "By the end after four cycles with the oligo pools, the overall success rate was about fifty percent. It was made possible by data that was impossible to collect beforehand."

We believe proteins designed completely from scratch will revolutionize biotechnology.

David Baker

"Large scale testing also enabled a really incredible thing, which was to test thousands of control sequences that were similar to the designs, but were expected *not* to fold," Rocklin added. "To generate similar sequences we expected to fail, we made mutations at key residues, or shuffled the order of the amino acids, which keeps the key physical properties of the protein molecule identical. One would never test controls like this when testing designs one at a time — it's too much work. But it's easy in a large-scale experiment. These controls showed us that many of our designs were much, much more stable than closely-related control sequences, which gave us confidence that the designs were stable because they folded *as designed*, rather than folding into some serendipitous, non-designed structure."

Large-Scale Design for Therapeutic Candidates

In the other study, a team led by Rocklin and Aaron Chevalier and Daniel-Adriano Silva, postdoctoral fellows at Baker Lab and the University of Washington, again tested the protein folding process and also designed proteins to bind to two different targets: virulence factors causing influenza and botulism. Since the proteins were designed to bind to disease-causing targets, they are considered potential therapeutics. By binding to the targets, the molecules could block the virulence factor's disease causing activity. The team produced proteins that target influenza haemagglutinin and botulinum neurotoxin B, along with control sequences to probe contributions to folding and binding, and identified high-affinity binders. By comparing the binding and non-binding design sets, which are two orders of magnitude larger than any previously investigated, the team was able to evaluate and improve their computational model, leading to an increase in design success.

Designing a protein that folds is arguably a simpler challenge than designing a protein that folds into a structure that can bind to a specific target molecule. However, cracking this problem opens up the potential to treat serious global disease threats. Again, the ability to use large amounts of oligonucleotides at competitive prices allowed the researchers to use the power of scale to circumvent this challenge. "Because of the difficulty of that task, if we were testing tens of designs at a time, we might not have gotten anything," Rocklin said.

"In both papers, we took some of the successful designs — whether they were designs that folded properly, or designs that bound to the target — and solved structures of a few of them," Rocklin explained. "In both papers we solved the structures of a small subset of the protein panel. Those structures, along with the thousands of control sequences, gave us confidence that the thousands of other successes that we got were also likely folded as designed, even though we didn't structurally characterize every one of them."

Creating Computationally Designed Scaffolds

"There is a lot of interest in computationally designed non-antibody scaffolds, as they can be much more stable than scaffolds that people find from natural proteins," Rocklin noted. "Because of the stability of computationally designed proteins, you might ultimately have more robust drug candidates that can be stored more easily. That paper and others have shown that you can get similar binding activity to an antibody with computationally designed proteins."

In both experiments, Rocklin and his colleagues were able to create new, folded protein designs that could direct future therapeutic and synthetic protein development. Twist Oligo Pools enabled their high-throughput experiments with the flexibility needed to move forward.

"Our ability to get huge numbers of oligos affordably and also test them all in pools, made the experiments possible," Rocklin said. ■

WHAT CAN TWIST DO FOR YOU?

sales@twistbioscience.com
twistbioscience.com

#WeMakeDNA