

TECHNOLOGY FEATURE

THE COMPUTATIONAL PROTEIN DESIGNERS

A new breed of protein engineer is finding that the best way to create a molecule is to build it from scratch.

BRIAN DALBALCON



By designing a protein from the ground up, researchers can create molecules with forms and functions not found in nature.

BY JEFFREY M. PERKEL

Cassie Bryan's success at crafting a protein that worked as she intended was a long time coming. When it finally happened, after six long years, she hit the bar and celebrated with beers — and a karaoke rendition of Joan Jett's 'Bad Reputation'.

Bryan joined the protein-design laboratory of David Baker in 2012 as a graduate student at the University of Washington, Seattle. Her project was to design a protein that could bind to PD-1 — a protein on the surface of white blood cells that throttles the activity of the immune system.

At first, Bryan did what protein engineers have long done: she tweaked an existing natural

protein to make it bind to PD-1. But, two years into her project, she decided that approach was going nowhere. And an explosion of interest in PD-1 as a cancer-immunotherapy target during that time meant her goalposts kept moving. Meanwhile, the lab was growing ever more adept at a different approach. Instead of modifying natural proteins to fit a particular need, the Baker lab began creating proteins from scratch.

Although considerably harder than conventional protein engineering, *de novo* protein design offers several advantages, says Brian Kuhlman, a protein engineer at the University of North Carolina at Chapel Hill, who as a post-doc with Baker in 2003 led the lab's first *de novo* success¹, a 93-amino-acid molecule called

Top7. Natural proteins are difficult to modify without disrupting their overall structure. But by making proteins from scratch, researchers can design proteins to be more forgiving. They can build enzymes with activities unknown to nature, using co-factors and amino acids that are not part of the standard macromolecular toolkit. And scientists can test their understanding of protein biology, to ensure that they truly grasp the fundamentals.

"We're making everything up from scratch," says Baker. "And that's a very strict rule in the lab: you're not allowed to start with anything that exists in nature, because we wanted to be able to be sure we understand everything and design everything from first principles." ►

► For the most part, these artificial proteins have been what Baker calls “rocks” — ultra-stable proteins, such as Top7, of defined shape that other researchers can build on. Over the past few years, however, scientists have grown ever more skilled at imparting function, creating everything from fluorescent and cell-signalling proteins to candidate vaccines. But they’re in the minority in the design community — Baker estimates that 95–99% of protein engineering “is still done by random mutation and selection”. And *de novo* protein engineering often requires weeks of computational time and months of iteration. Still, computational advances and a broadening user base is making the process more accessible.

“It’s a tremendous time to be in this area,” says Donald Hilvert, a protein chemist at the Swiss Federal Institute of Technology (ETH) Zurich, who has worked with Kuhlman to create enzymes called esterases. “The combination of computation, structure, molecular biology, detailed biophysical measurements — all of this is coming together in such a beautiful way.”

IT'S COMPLICATED

Protein folding is complicated. Built as long chains of amino acids, newly formed proteins quickly collapse into a specific folded shape, from which the molecules derive their function. Researchers have long known that a protein’s sequence defines its shape. And they can experimentally determine that shape using X-ray crystallography and cryo-electron microscopy. What they could not do was predict the shape from the sequence alone.

That’s because a protein’s structure is defined by multiple competing forces. A protein is basically a long string of carbon, nitrogen, oxygen and hydrogen, with amino-acid side chains dangling like charms on a molecular bracelet. The molecule cannot fold into just any shape, however — the possibilities are constrained as different parts of the protein jostle for position and balance attractive and repulsive forces. The trick in protein-folding prediction is to work out those forces, and thus the precise angles that the protein bonds will take.

The Baker lab uses a suite of molecular modelling and search tools called Rosetta, which can calculate the energy of a folded protein and search for the lowest energy sequence for a given structure, or the lowest energy structure for a given sequence. Baker developed Rosetta in the late 1990s as a tool for predicting structure. The software has been under continuous development ever since, both by members of his lab and a community of several hundred users called the Rosetta Commons, to improve its performance and capabilities.

For instance, in a project to design short circular peptides called macrocycles — which can have antibiotic and anticancer properties — Baker lab postdocs Parisa Hosseinzadeh, Gaurav Bhardwaj and Vikram Mulligan (who is now at the Simons Foundation in New York City) collaborated² to teach Rosetta how to

handle ‘D’ amino acids. These are chemical mirror images of the ‘L’ residues used by cells, and therefore have different properties. Protein designer Neil King, a Baker lab alumnus who is still at the University of Washington, has modified Rosetta to design self-assembling protein nanoparticles.

Although each *de novo* project in his lab is different, Baker says that they all follow the same basic strategy. First, decide on a desired class of structures — a ‘Platonic ideal’ of a shape, as he puts it. Then, use Rosetta to design tens of thousands of potential backbone conformations to match that shape, flesh those out with side-chain residues, and test that the calculated sequences will fold into the desired form. Finally, synthesize genes that will express the best designs, test, iterate and repeat.

“Only a very small fraction of possible backbone conformations are actually designable,” Baker says. And researchers might need to search through millions of possibilities and dozens of physical proteins before selecting the right candidate. Zibo Chen, a graduate from the Baker lab who is now at the California Institute of Technology in Pasadena, sifted through some 87 million backbones to identify 2,251 designs that are capable of protein–protein interaction. The computation took about six weeks on several hundred processor cores.

Inspired by DNA origami — in which DNA molecules are folded into nanostructures — Chen wanted to identify hydrogen-bonding strategies that would allow him to design perfectly orthogonal protein pairs (proteins that would interact only with a specified artificial partner, but not with other similarly designed proteins). Such proteins could be used to create novel biosensors, genetic circuits or just whimsical shapes. Chen

joined the lab, he says, partly because he wanted to see whether he could recreate with protein what DNA nanotechnologists had made with nucleic acids: a macromolecular smiley face emoji. Earlier this year, Chen described the first step towards such a design: a self-assembling 2D array³. “I was quite naive about what I could achieve in five years,” he says.

Bryan designed her protein — all 46 amino acids of it, tiny by protein standards — to interface with, and hopefully regulate, PD-1. The protein, she says, is simply a flat surface — a β -sheet — scaffolded by a single, rod-like α -helix. In cartoon form, it resembles an old-fashioned iron used to press clothes. “The helix is kind of like a handle, and the actual functional end is the iron that sticks to the receptor,” she explains.

Bryan first tried to modify an existing protein to assume that shape, but found she couldn’t produce the protein in a usable form. So, inspired by the known structure of PD-1 binding to its natural ligand PD-L2, she identified three crucial residues, coded their positions

into Rosetta and directed the software to build a protein that would support it. She extended an essential loop by five amino acids to improve binding to the human target. And using a high-throughput screening strategy based on flow cytometry (a cell-analysis technique) and DNA sequencing, she tested every amino-acid variant at every position to nudge the structure towards ever-stronger interactions. On the way to designing her protein, Bryan received her degree, despite a three-year detour when she realized that her engineered protein couldn’t interact with its human counterpart owing to some crucial sugar modifications.

Finally, Bryan had a breakthrough: the protein bound to lymphocytes in a flow cytometer. With so many ups and downs, Bryan was sceptical of reading too much into any one experiment, she says. But those flow data, provided by her immunology colleagues, made her believe. “It was these immunology collaborators who know T cells really well, and they’re telling me that on real human T cells from real people, we saw this strong effect that hadn’t really been seen before with similar molecules.”

King, who has designed a self-assembling nanoparticle that could serve as a candidate vaccine for respiratory syncytial virus⁴, describes shepherding a molecule from concept to reality as surreal. “You’re making it up,” he says. “It’s literally a computer fantasy. And when it actually works in the real world, it’s just magical.”

And so Bryan celebrated, as she says, with beers and Joan Jett.

DESIGNING FOR FUNCTION

At this point, there’s little that protein engineers cannot do, Baker says — at least in terms of shape. But most proteins don’t exist simply to assume a specific shape; it’s function that matters.

Function, such as the ability to catalyse a chemical reaction, complicates design, says Hosseinzadeh, because it adds new variables to the problem. “When I pick for shape, the only thing I care about is the overall energy,” she says. “But when you design for function, there are certain other things that come into consideration — for example, does this molecule make good contacts with the protein surface that I want to target? Are the targeting side chains positioned in the correct place? And does it cover the [interaction] surface?”

When Anastassia Vorobieva, a postdoc in Baker’s lab and Jiayi Dou, who is now at Stanford University in California, decided to create a *de novo* analogue of green fluorescent protein, the two researchers came to the project with different agendas. Vorobieva wanted to create a β -barrel, a common structural motif that had yet to be created from scratch; Dou wanted to build a protein that could stabilize a small molecule, such as a fluorophore.

A β -barrel is a structure in which one edge of a β -sheet connects with the other, creating a hollow pore or pocket. But they are particularly tricky to create, Vorobieva says, because the





Cassie Bryan at the University of Washington built a protein that binds to the cell-surface protein PD-1.

BRIAN DALBALCON

individual threads of the sheet are sticky; if the protein isn't designed just so, it will degrade into a useless mess.

Vorobieva's aim was to create a barrel with a smoothly curving surface. But that design placed an unexpected strain on the peptide backbone. A few well-placed glycine residues imparted a squarish cross section, but relieved the stress enough for the design to succeed. Vorobieva showed this with a crystal structure that closely matched her concept⁵. That "was the final strongest experiment that showed we were doing everything right", she says.

To make the protein functional, Dou reproduced Vorobieva's original design, but with additional constraints to stabilize a fluorescent molecule. She worked with Baker lab research scientist Will Sheffler, who was designing a new Rosetta module to sample the possible binding conformations of a small molecule bound to a protein. Dou balanced stability and function by deliberately restricting the fluorophore to the top of the barrel. Dou identified 2,102 candidate designs, and synthesized 56. Two fluoresced in the presence of the fluorescent substrate, one of which Dou further modified to maximize brightness and validate her design — an effort that involved testing some 2,090 gene variants.

Protein design almost always involves selection and iteration, notes Lynne Regan, a protein chemist at the University of Edinburgh, UK. Researchers cannot yet sit down at a computer and design a protein that binds another molecule and get it right first time; they have to make something that works to some degree, and then improve on it.

In part, that's because researchers are still working out the minutiae of protein folding. Baker notes, for instance, that Rosetta depends

on its 'energy function', a model that estimates the energy associated with each structure. But just because the program says a molecule will assume a particular shape doesn't mean it actually will. Sharon Guffy, a protein scientist at biotechnology company Pairwise in Durham, North Carolina, who did her graduate work with Kuhlman, says she struggled to get Rosetta to correctly account for the electrical properties of zinc (and its impact on nearby side chains) when creating a metal-binding protein. "It cost me at least a month or so" of coding and troubleshooting, she says.

At the University of California, San Francisco, Marco Mravic, a graduate student in the laboratory of protein engineer William DeGrado, focuses his research on membrane proteins — specifically, their assembly into larger complexes. He chose to study a cardiac protein called phospholamban, which comprises five identical membrane-spanning helices. What is it, Mravic wanted to know, that directs these helices to assemble so precisely?

Part of the problem was structural. Nobody actually knew what phospholamban looked like. Mravic ran a molecular-dynamics simulation of the protein, which suggested the complex splays open at one end like a banana peel. "It was like, this simulation doesn't look right," Mravic says. "So I just went into the molecule and 'fixed' it."

By changing two water-loving amino acids to more membrane-favourable residues, Mravic created a more tightly packed variant, which he demonstrated by solving the crystal structure. He then worked out the features that allowed that packing to occur, identifying what he calls a "steric code" — a configuration of four amino acids on the helix surface that allow key side

chains to interlace like a zip. Mravic then used that code to design synthetic derivatives that adopt structures analogous to phospholamban⁶.

STRUCTURAL FOUNDATIONS

Beyond the nuances of protein folding, *de novo* design allows researchers to push the boundaries of what proteins can do. At the University of Birmingham, UK, for instance, chemist Anna Peacock studies metallopeptides — miniature proteins that bind metal ions. In biology, such molecules typically bind zinc, manganese or copper — "things that are found dissolved in seawater", she says. But other metals could enable different chemistry.

Peacock has used *de novo* proteins as scaffolds to create molecules capable of binding gadolinium, complexes of which are commonly used as contrast agents for magnetic resonance imaging. She is also crafting enzymes that can use metals such as platinum or iridium to explore reactions not found in nature. "I don't personally see the point in getting an artificial metalloprotein to do the same chemistry that an enzyme can already do," she says.

As each design goal is achieved, it becomes easier for others to emulate them. The Baker lab has even developed an online gaming interface to Rosetta, called FoldIt, that challenges players (few of whom are scientists) to create proteins *in silico*. In a study this year analysing their work⁷, the players delivered. They built novel designs "completely from scratch", Baker says, including one fold that had never been seen before.

Few scientists have the time or expertise to design a protein from the ground up, of course; for them, *de novo* designs are foundations to build upon. But in the Baker lab, the design work continues. With each success, the lab celebrates. For the postdocs and students who do the work, Baker says, the euphoria "lasts for quite a long time. For me, it lasts for a day or two, and then it wears off and I'm like, okay, what are we gonna do next?" ■

Jeffrey M. Perkel is technology editor at Nature.

1. Kuhlman, B. *et al. Science* **302**, 1364–1368 (2003).
2. Hosseinzadeh, P. *et al. Science* **358**, 1461–1466 (2017).
3. Chen, Z. *et al. J. Am. Chem. Soc.* **141**, 8891–8895 (2019).
4. Marcandalli, J. *et al. Cell* **176**, 1420–1431 (2019).
5. Dou, J. *et al. Nature* **561**, 485–491 (2018).
6. Mravic, M. *et al. Science* **363**, 1418–1423 (2019).
7. Koepnick, B. *et al. Nature* **570**, 390–394 (2019).

CORRECTION

The Toolbox article 'Drone takes to the skies to image offshore reefs' (*Nature* **570**, 545; 2019) gave the wrong affiliation for Ved Chirayath. He is director of the NASA Ames Laboratory for Advanced Sensing in Mountain View, California. Also, the picture caption erroneously stated that the drone was carrying a 'fluid cam'. In fact, it is a commercial camera.