



Credit: Institute for Protein Design

The hothouse for protein design

The University of Washington's Institute for Protein Design has become a hub to galvanize de novo protein engineering, citizen science and much more since its founding in 2012.

Caroline Seydel

Researchers at the University of Washington's Institute for Protein Design (IPD) don't like to work alone. They see the power of citizen science in galvanizing their internal projects. In March, the IPD announced the winners of its first Coronavirus Spike Protein Binder puzzle, which challenged players from around the world to design molecules that could potentially block SARS-COV-2 infection. This first of what will be three such challenges (Box 1) generated a diverse array of potential binders, 99 of which are currently being synthesized for testing at IPD. At the center of this is University of Washington professor David Baker, who founded IPD in 2012 and has shepherded an ever-expanding ecosystem

of collaborations between IPD and other academic labs, biotech startups (Fig. 1) and citizen scientists. The institute has garnered recognition outside the usual academic setting; last year, it won a five-year, \$45 million grant from the Audacious Project, a funding initiative at TED, which will allow them to double their faculty while creating a "Bell Labs of protein design," a playground of invention for scientists with diverse backgrounds and skill sets. The IPD actively promotes collaboration with other institutions, enhancing their creative and technical expertise.

If you ask Baker what excites him about the institute, he doesn't mention grant money or hiring sprees. "It's exciting that we're now getting to the stage where we

can build new things that can be useful in the world," he says. Baker believes that de novo protein design technology has turned a corner, from creating functional but simple proteins to building complex molecular machines that can perform conditional operations. No longer limited to describing and copying nature's proteins, protein designers have learned to build sets of proteins that can change conformation on demand or carry out multi-step instructions, such as toggling between two states.

Today IPD lists seven ongoing projects to address the pandemic, from designing nanoparticle vaccines and anti-inflammatory proteins to screening existing drugs. Through Baker's hypercollaborative nature and his desire to distribute the technology

Box 1 | Protein designer citizens unite!

To perform its complex calculations, Rosetta requires a massive amount of computing resources. To access more processing power, the Baker lab in 2005 launched *Rosetta@home*, a distributed computing network that recruits volunteers to run Rosetta in the background on their home computers. Since then, some 100,000 host computers in over 150 countries have joined the network.

The project attracts casual users as well as hard-core ‘overclockers’, gear enthusiasts who put together custom computer systems to maximize their processing power. These tech hobbyists build extensive processing and custom cooling arrangements to push the limits of their hardware. Overclocker

clubs sometimes host competitions to see whose rig can process the most data in a weekend, and some clubs are sponsored by computer hardware manufacturers.

In February, Rosetta successfully predicted the 3D structure of the SARS-CoV-2 spike protein, which allows the virus to gain access to human cells. The flurry of effort that accompanied the project crashed some of the networked computers, but with the protein structure now in hand, work is underway to fight the disease. Researchers crafted mini-protein binders that perfectly complement the spike protein, preventing it from latching onto cellular receptors. To make a vaccine, IPD researchers are attaching the spike

protein to a synthetic virus-like particle. The particle displays the spike protein in a repetitive array, an arrangement designed to stimulate a vigorous immune response.

Rosetta fans participating from home aren’t limited to contributing electronic processing power. Protein design requires human ingenuity as well, and for that, there’s the *Foldit* game. With a video-game-style interface, *Foldit* invites users to shift protein backbone structures and rearrange side chains to achieve the most favorable free energy configuration. *Foldit* players interact as part of a cooperative community, discussing problems in the forums and working in teams to solve structures.

for broad adoption, de novo protein design may one day become a part of every protein engineer’s toolbox.

Unnatural solutions

At the heart of the IPD is the idea that modern problems require modern solutions, designed by humans. Mother Nature has created some incredible tools, but those solutions have come about as a result of evolutionary pressures, including the need to conserve genetic real estate by making proteins that serve multiple functions. Whereas a natural protein may be the solution that evolution has arrived at, it may well not be the most efficient tool for a task thought up by a human. Creating new proteins, Baker says, allows the tools to be specifically

directed, as well as modular and customizable to other uses.

The second key to the IPD is that it draws on generations of work in biological sciences and protein engineering, funneling that vast reservoir of observational data into designing and building new proteins from scratch. “It is very much the case that the work that can be done in protein design is always going to depend on work that is being done outside the field of protein design,” says Ian Haydon, a former student in the Baker lab who now serves as scientific communications manager for the IPD. “We have an interesting and valuable suite of technologies here, but it’s worthless without partners who are specialists in other areas.” Those specialists, he says, provide deep understanding of the biological

underpinnings of disease or other cellular processes, to help direct the application of protein design technology.

“People have described these marvelous biological systems for a long time, but it’s always, hey, look at these cool things we found in nature,” says Baker. Observing and experimenting provide explanations for complex biological systems, but creating a system from scratch requires a deep understanding of how all the parts work together. “When you build it, you really have to think about what the principles are.”

Before there was an IPD, there was Rosetta. The original concept Baker and colleagues had in mind was to develop algorithms to predict a protein’s final 3D structure. Rosetta simulates the interplay of hydrogen bonds and side chain attractions

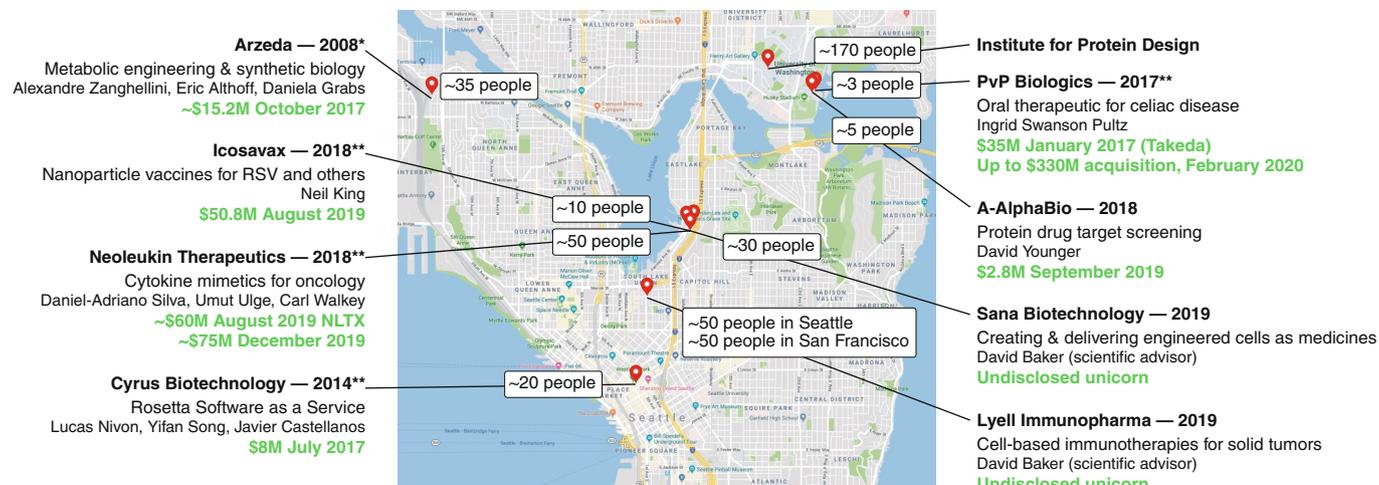
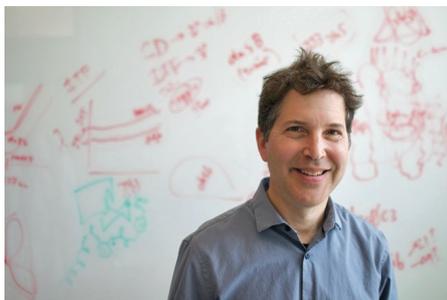


Fig. 1 | Spinouts from the Baker lab and IPD Translational Research Program. Since before there was an IPD, Baker has been spinning out companies. At last count, there are eight. *Baker lab spin-out, pre-IPD. **IPD Translational Research Program spinout.



David Baker, director of the Institute for Protein Design. Credit: Institute for Protein Design

and repulsions through which a linear chain of amino acids collapses into a thermodynamically favorable state. In the early 2000s, in work spearheaded by then-postdoc Brian Kuhlman (now at the University of North Carolina, Chapel Hill), Baker's team figured out how to use Rosetta 'backward'. In addition to predicting the folding conformation of a particular sequence, they taught the computer to predict the amino acid sequence that would form a given shape.

Over the years, the program has been expanded and updated by a collaborative community of developers around the world, called Rosetta Commons. "Before David, the field was pretty small, and we all had our own software that we were developing," says William DeGrado, one of the pioneers of *de novo* protein engineering at the University of California, San Francisco (UCSF). "Rosetta has been just a tremendous accomplishment—not just the science, but building the community. That has brought in so many amazing collaborations."

Humans have ideas; Rosetta gives them shape

Rosetta is far more efficient than human labor at finding amino acid sequences that will collapse into a desired structure, but making the protein perform a desired task ramps up the complexity. For that, there's no substitute for human ingenuity.

As a grad student in Samuel Miller's lab at the University of Washington, Ingrid Swanson Pultz was studying bacterial communication, not designing novel proteins. She drifted into Baker's orbit when she formed a team at the University of Washington for the 2011 International Genetically Engineered Machines (iGEM) competition. "The team worked really closely with David's lab," she says. Their goal: to make a therapy for celiac disease by designing an enzyme that could break down gluten.

In people with celiac disease, the body mounts an inflammatory response to one

particular peptide in the gluten protein. Enzymes exist that can break down the offending peptide, but they don't survive in the harsh, acidic stomach environment. The team started with a bacterial enzyme hardy enough to withstand the trip through the human digestive tract and set out to make it act on the immunogenic region of gluten. Using Rosetta and particularly the Foldit game, the team rearranged the amino acids into an active site that would break down gluten.

"The students generated hundreds of different designs," says Pultz. After synthesizing a hundred or so candidates, they identified some that actually worked. "That year we went to the iGEM competition, and the UW team actually won the competition, which was the first US team ever to win the award."

After the students went back to their own work, Pultz didn't want to let the project languish, so she continued to work on optimizing the enzyme until she had something that could break down a meal-sized serving of gluten in a physiologically relevant time frame. In 2015, she launched a company, PVP Biologics, based on the technology. PVP was the first company launched through the University of Washington's Translational Investigator Program. Takeda Pharmaceuticals fronted the money to conduct phase 1 clinical trials on the novel enzyme, called KumaMax, and in February 2020 exercised its option to acquire PVP.

Eliminating side effects

KumaMax successfully adapted an existing enzyme to perform a new function, but in some cases, success requires assembling available parts into something completely new. "One of the promises of protein design is that you can design and build brand new proteins to solve current-day problems for which we don't already have evolved proteins to solve," Baker says. *De novo* protein design lets the designer select just the sequences and structures they want and leave out anything they don't. "They don't carry along that baggage as the result of a long evolutionary process," says Haydon.

The synthetic cytokine neoleukin-2/15 was created in an attempt to jettison some very heavy baggage that has dogged cancer researchers for decades. Since the 1980s, the cytokine interleukin (IL)-2 has tantalized oncologists with the promise of harnessing the immune system to attack tumors. IL-2 naturally stimulates T cell proliferation, but it also exerts ferocious side effects on the body, particularly in the lungs.

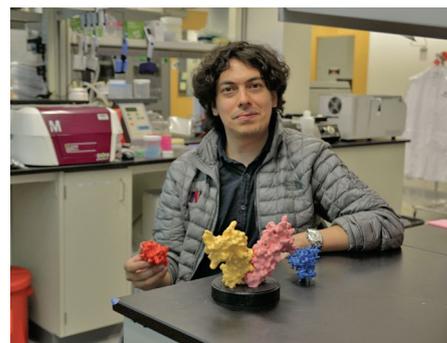
Some effector T cells express a version of IL-2 receptor containing three subunits,

alpha, beta and gamma, one of which, alpha, suppresses the immune response, and mediates some side effects of IL-2 that are toxic to the lungs. IL-2 binds cells expressing the alpha subunit with a much higher affinity than the effector T cells that lack it, so the quest has been to make a modified form of IL-2 that doesn't bind the alpha subunit. This should reduce toxicity, while focusing the protein on those effector T cells that will go after the cancer.

By and large these efforts had been unsuccessful, partly because the modified IL-2 proteins were too unstable for manufacturing and storage. Daniel-Adriano Silva, who is now the vice president head of research at Neoleukin Therapeutics, recalls tackling this problem during his time in the Baker lab. "I was drawn to this project as one of the challenges that people had tried for several years already that didn't work, to see whether I could apply this technology to take inspiration from the natural IL-2 to try to build this antagonist," he says.

Silva studied the 3D structure of IL-2 in its high-affinity conformation, the shape it assumes when bound to all three receptor subunits. He built a new protein that emulated that shape but was "otherwise unrelated in topology or amino acid sequence." It also lacked any binding site for the alpha receptor subunit. "The trick here is that we are not building IL-2," says Silva. By constructing a completely novel protein instead, the team could create a highly stable molecule that bound with high efficiency to the beta and gamma subunits and ignored the alpha subunit.

Once the concept was in place, the actual building process required a lot of testing and tweaking of prototypes. "What was integral was having the collaborative back and forth," says biomedical engineer Jamie Spangler, of Johns Hopkins University. She worked on the neoleukin project while she was a postdoc in Christopher Garcia's



Daniel-Adriano Silva, vice president and head of research at Neoleukin Therapeutics. Credit: Neoleukin Therapeutics

lab at Stanford University. The Baker lab designed candidate proteins using Rosetta, and once they were expressed and purified, they sent them to Spangler to test how well they bound to the receptor. Although the computational simulations were highly accurate in predicting protein structure, they couldn't predict how well the protein would interact with its receptor. "Ultimately, the neoleukin-2 molecule that was published was a much later generation of those initial formulations and designs," she says. "It took both the computational and the experimental evolutionary workflows to ultimately settle on that really effective molecule."

Although native IL-2 binds preferentially to receptors containing all three subunits, neoleukin-2/15 exhibits the same affinity for receptors with or without the alpha component. In mice, the protein successfully boosted the proportion of cytotoxic T cells in tumors without inciting the adverse effects seen with IL-2 itself. It's also hyperstable, retaining its activity even after exposure to 95 °C temperatures, making storage and manufacturing for mass distribution a plausible goal. Silva, Baker and their colleagues launched Neoleukin Therapeutics in January 2019 to develop the protein into a commercial cancer therapy. That August, the company merged with Vancouver-based Aquinox Pharmaceuticals to become a publicly traded entity, the first IPD spinout to go public, under the name Neoleukin Therapeutics.

Template-free designing

Protein design techniques have performed best when called upon to create regular, ordered structures, but functional proteins often rely on irregular motifs to do their jobs. Working from a template, as was done in designing neoleukin, can help researchers fashion functional motifs, but also imposes certain structural limitations on the final product, and it can require a lengthy optimization process in vitro to get to the desired end product. Bruno Correia, a former student in the Baker lab who now runs his own lab at the Ecole Polytechnique Fédérale de Lausanne, Switzerland, recently published TopoBuilder, a template-free design algorithm based on Rosetta. Correia's group used TopoBuilder to design molecules with optimized topology for stabilizing antigenic epitopes

TopoBuilder allows the designer to take the functional elements into account from the start and custom-build the scaffold around those elements, rather than optimizing the overall protein after the fact. Correia used TopoBuilder to make proteins that present respiratory syncytial virus

(RSV) F-antigen epitopes. "We are tailoring the protein scaffolds into the epitopes that we need to present to the immune system," says Correia. The scaffolds created with TopoBuilder present three viral epitopes and successfully induce antibody production in mice. Still, Correia cautions, the paper is a successful demonstration of the TopoBuilder approach to creating a particular structure, but has not proven itself in terms of vaccine design.

Another vaccine design strategy produced at the IPD is considerably farther along in development. Neil King, a former Baker lab postdoc who now has his own lab at the IPD, spun out the company Icosavax in 2018 based on his self-assembling nanoparticle technology. King created protein subunits that arrange themselves into a virus-like particles (VLPs), displaying viral antigens in regular arrays to elicit an effective immune reaction. Synthetic VLPs offer more opportunities for customization than natural ones, opening up the possibility of vaccines for intractable viruses such as RSV and HIV. King and his colleagues fused the SARS-CoV-2 spike protein to self-assembling nanoparticles and are working on creating the optimal antigen display for a coronavirus vaccine.

Going small

Protein subunits that self-assemble into large VLPs show how protein design can achieve increasingly complex structures. On the other end of the spectrum, designer peptides capitalize on the advantages of small size, combining the stability of a small molecule with the specificity of a large antibody-based drug.

Gaurav Bhardwaj, a former postdoc in the Baker lab who now heads his own lab at the IPD, has developed computational tools to design peptides incorporating unnatural amino acids. This approach opens up a whole new toolbox to create molecules with unique properties. "We should be able to build more chemically and structurally diverse peptides because we have a larger building set," Bhardwaj explains. "But to use these non-canonical amino acids, we had to build a lot of new computation and experimental methods to work with these amino acids."

Unlike antibodies, designer peptides can be endowed with the ability to cross the cell membrane; like antibodies, they bind tightly and specifically to their targets. Building peptides with non-canonical amino acids that have reverse chirality also helps them avoid being attacked by natural enzymes, which evolved to recognize natural amino acids. This could help make drugs that can be given orally because they can survive the harsh digestive enzymes.

"It's still beginning stages," said Bhardwaj. "It's a very exciting time, because there's so much happening in this space."

Producing sustainable materials

Protein design has plenty to offer outside the medical realm as well. Arzeda, an IPD spinout that bills itself as "The Protein Design Company," uses custom-designed enzymes to brew specialty chemicals and healthy food ingredients in large vats of bacteria. They have created biochemical pathways for various industrial partners to produce specialty chemicals sustainably, such as a durable, scratch-resistant bioplastic for cell phone screens based on a compound found in tulips. Their mission is to use protein design technology to engineer sustainable sources of chemicals and ingredients that not only replace petrochemicals, but outperform existing products.

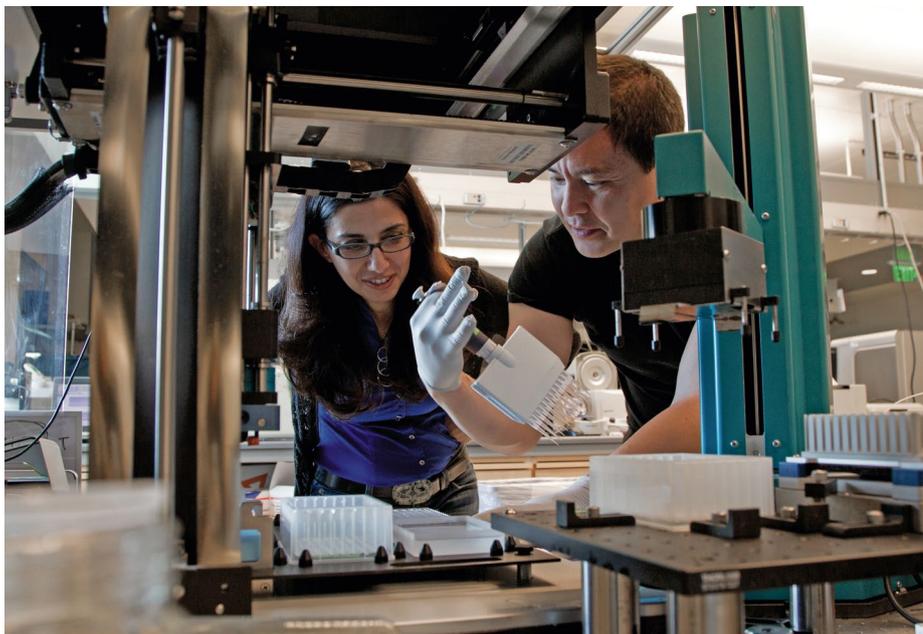
"IPD is really focused on pushing the envelope, developing new algorithms and new concepts that you can use with protein design," says Arzeda co-founder Alexandre Zanghellini, a former graduate student in the Baker lab. "We are focused more on how you take that technology and scale it up and automate it in a way that you get improvement in performance and decrease in cost."

In December 2019, Arzeda renewed a collaboration with BP that began in 2018 to develop biological manufacturing pathways for industrial chemicals. "By using biology instead of synthetic chemistry, you can actually create molecules that would not be accessible by synthetic chemistry and that have better properties than an oil-based chemical," says Zanghellini. "We are using biology to make something much better compared to what oil would make."

Proteins in a box

Designer enzymes open up a range of possibilities, but ultimately each enzyme performs one specific task. A living organism consists of countless complex interacting signaling pathways and feedback systems, working together to maintain homeostasis. To recreate such a system, protein designers needed to find a way to make proteins with two stable conformations.

As a systems biologist, Hana El-Samad of UCSF, didn't expect to find much overlap between the feedback control systems she was developing and Baker's atom-by-atom computational protein design. But when the two happened to have a chat, they soon realized that Baker's newest project, a set of proteins designed to change conformation in the presence of a molecular 'key', could



Hana El-Samad (left) works with postdoctoral scholar Ignacio Zuleta on a machine that measures instantaneous growth and gene expression dynamics. Credit: Susan Merrell

be customized to build a protein-based feedback control system. “It was truly serendipity,” she recalls. “It was two trains of thought that had a very low chance of colliding, and they just did.”

The protein system is called LOCKR, for “Latching, Orthogonal Cage/Key Proteins,” and it consists of a six-helix protein that folds into a stable, cage-like formation. The cage can interact with either its own ‘latch’ helix domain, which holds the cage closed, or to another protein—the ‘key’—which triggers a conformational change. When the key comes along, the protein releases the latch, which flips open like a switchblade. By encoding a bioactive peptide into the latch, the LOCKR system can be programmed to perform a specific function only when the active domain is unveiled in the presence of the key.

Designing such a system from scratch requires the designer to come up with proteins that have multiple configurations that are thermodynamically favorable. If the cage binds the latch too loosely, it will flip open spontaneously, and if it binds too tightly, the key won’t be able to force it open. Precise control depends on achieving a balance between the two forms. “From the idea of installing function to final design took a lot of iterations,” said Bobby Langan, a former graduate student in the Baker lab and one of LOCKR’s designers. “In the early designs, it would be ‘leaky,’ and there would be some activity in the off state.”

Langan and his colleagues were testing the system by embedding a Bim peptide into the latch domain. In the presence of the key, the LOCKR would expose Bim and initiate binding to its target, Bcl2. “We were not using it for its biological function,” Langan pointed out. “But we were able to cage that protein–protein interaction, and that was the jumping off point.”

The beauty of the LOCKR system is that it’s modular and customizable, so rather than going back to the drawing board for each new cellular application, it’s possible to swap out the active domain with different functional elements. “David said, it opens and closes, and I said, well, why?” recalls El-Samad. “What did you put inside?” The pair discussed the idea of enclosing a degenon, a protein sequence that helps regulate protein degradation.

El-Samad’s return to San Francisco commenced “an intense and beautiful” collaboration, in which Baker’s team would send her computationally derived protein sequences and her lab would synthesize the proteins and test them. “It was very clear from the get-go that thing was going to work,” recalls El-Samad. Over the course of several months of testing and optimizing, the teams developed two more inducible LOCKR systems. The first one, degenLOCKR, encodes a cODC degenon, and expression of the key initiates its degradation. The second, nesLOCKR,

encodes a nuclear export sequence, and the key successfully induces its removal to the cytoplasm. LOCKR activation is tunable to different concentrations of the key protein, making the technology adaptable to a variety of cellular activities.

In addition to inducible protein switches, Baker’s group this year published a heterodimer-based system for conditional activity: protein logic gates. Logic gates are preset systems that produce a specific response depending on the particular set of conditions they encounter. The simplest type, the AND gate, acts only when two signals are both present; the OR gate responds when either input signal is present, and so on. To build the conditional system, the team had to first develop stable orthogonal heterodimeric pairs, like a protein version of DNA base pairs. Then they linked heterodimer pairs to functional elements in such a way that the functional elements will only get together and form active complexes when the heterodimers hook up. To evaluate the synthetic proteins’ ability to pair up, the researchers turned to native mass spectrometry (Box 2).

In an April 3 *Science* paper, the Baker lab demonstrated a NOT gate that could overcome T cell exhaustion by repressing the immune checkpoint gene *TIM3*. Imagine two heterodimer pairs, A:A’ and B:B’. By genetically fusing monomer A to a DNA binding domain, TALE, and fusing monomer B’ to a repressor domain, KRAB, they established a mechanism for controlling gene expression. To bring the TALE and KRAB domains together and repress *TIM3* expression, they expressed another construct, containing subunits A’ and B, capable of simultaneously binding the A-TALE and the B’-KRAB fusion constructs. Only when the A’-B construct was absent could *TIM3* expression proceed.

These biological logic gates are an essential step toward building complex, ‘smart’ therapeutics. “Cells are always making complicated decisions based on multiple inputs,” says Baker. Smart therapeutics, he says, need to be able to calculate the appropriate actions inside the cell based on changing conditions: when is it the right time to activate a gene or kill the cell? To build in that decision-making capability requires a modular and customizable set of decision tools that can be deployed in various situations. “Nature’s solutions tend to be very elegant, but they’re always tied to a specific situation,” says Baker. “If you had a very general way of doing computations inside cells, then you could just plug that in.”

Box 2 | Analyze that!

Bioanalytical chemist Vicki Wysocki of Ohio State University joined forces with the Baker lab several years ago, when one of her students persuaded Baker to send over some samples to characterize with native mass spectrometry. Wysocki is the director of the National Resource for Native Mass Spectrometry-Guided Structural Biology at Ohio State, and her lab has pioneered a variety of mass spec techniques for characterizing protein structures and interactions.

“X-ray crystallography requires crystals, NMR [nuclear magnetic resonance] requires a lot of pure sample, cryo-EM [electron microscopy] requires getting the samples frozen properly—those techniques are labor intensive, and they don’t always work,” explains Wysocki. “Because the Baker lab makes so many complexes—for one project they might send a hundred or more samples—you can’t possibly make measurements with a technique like X-ray crystallography for all of those samples.”

Wysocki’s team [developed an automated workflow](#) allowing them to screen proteins in just three to five minutes per sample.



Vicki Wysocki (center), shown with (former) students Mengxuan Jia and Christine Wachnowsky, has developed enabling technologies for analyzing complexed proteins. Credit: The Ohio State University

Using native mass spec, the researchers could not only verify that the proteins that had been produced were as expected, but also characterize the protein–protein interactions within the mixture. “What’s really beneficial for this work in particular is that you can measure interactions between the proteins they designed,” says Zachary VanAernum,

a graduate student in the Wysocki lab. For instance, in the logic gate work, it was critical to understand which components were assembling into complexes, and with what frequency. For instance, consider two heterodimer pairs, A:A’ and B:B’. In a mixture containing monomer A, monomer B’, and a fusion construct of A’–B, they were able to detect that the components preferentially assembled into the three-part complex A:A’–B:B’ rather than either dimer A:A’–B or A’–B:B’.

For larger complexes involving multiple different subunits, the team was able to detect the configuration of the proteins in the complex using a mass spec technique called surface-induced dissociation. By slamming the molecules into a surface to break apart the subunits and then measuring which subunits colocalized among the resulting fragments, they could see how the subunits were arranged in the large complex. This type of structural characterization provides critical information to verify whether synthetic protein complexes are forming the way the designers expect.

Designing protein for the masses

As the pace of invention accelerates inside the IPD, the protein design revolution may be ready to spill over into more and more labs. In an April 2020 publication, the Baker lab reported [a design method using modular components](#) that could simplify de novo protein design for labs without the powerful computing resources of the IPD. “It introduces a lot of different tools that make the protein design faster and more efficient, and more likely to produce what you want,” says TJ Brunette, a research scientist in the Baker lab. “It can take weeks to design a single protein from scratch,” he says. These modular design tools speed up the process by providing defined elements that can be

joined together to form more complex shapes. “Let’s say you just want to orient cell receptors: you could use these proteins in the software to just click things together or trace the protein into whatever shape,” Brunette says.

Even as the Baker lab makes more protein design tools freely available, the field remains highly specialized. “It’s not like somebody can just walk into a room and start coding in Rosetta,” points out Jamie Spangler. “You have to have a fundamental understanding of the computational tools, but beyond that you have to have a fundamental understand of the biology behind it.” Beyond access to the tools, having access to a team with diverse expertise is key to succeeding

at de novo protein design. As interdisciplinary centers gain popularity, more labs may be able to take protein design in new directions.

Despite an impressive list of accomplishments, there’s a clear sense of anticipation for what’s coming next for the IPD. “We can only design a very small portion of structures, and we’re just finding really cool things to do with that small portion,” Brunette says. “It’s all very much at the beginning stages.” □

Caroline Seydel
Los Angeles, CA, USA.

Published online: 23 June 2020
<https://doi.org/10.1038/s41587-020-0586-0>