Putting Proteins Under a Computational Microscope

Grand Prize winner Zhuoran Qiao uses generative AI to potentially radically reform drug discovery.

Sometimes, even award-winning scientists struggle with lab work.

As an undergraduate at Peking University, Zhuoran Qiao dreamed of specializing in total synthesis, using sophisticated lab methods to create complex organic compounds. Then he got a wakeup call. “It's like being bad at cooking—I’m just not super-gifted when it comes to following synthesis workflows,” Qiao laughs. “I kept messing up and hurting myself.”

After several scorched fingers and failed experiments, Qiao changed track. “It turns out I’m reasonably good at using computers to model chemical reactions,” he says. “Working *in silico* tends to go better for me.”

That's an understatement. This spring, Qiao received the inaugural Chen Institute and *Science* Prize for AI Accelerated Research for his work using generative AI to model protein structures—a breakthrough that could transform pharmaceutical research.

“Zhuoran’s work is a stunning demonstration of AI’s power to accelerate discovery—thanks to his work, research that previously required years of months of laborious work can be completed in seconds,” says TCCI co-founder Chrissy Chen. “By empowering researchers to rapidly design new molecules, his work opens the door to important new treatment pathways for previously ‘un-druggable’ diseases.”

The road to discovery

After studying computational chemistry in China, Qiao headed to Caltech in 2018 to complete his PhD in the labs of Prof. Anima Anandkumar and Thomas Miller, a pioneering chemist and AI innovator. There, Qiao began using machine learning to bridge between quantum data and higher-level descriptions of molecular structure. “Using AI, we could obtain structural insights that previously required a prohibitive amount of computation,” Qiao says.

In 2021, as Qiao was refining his model, Google DeepMind underscored the potential power of computational modeling by unveiling AlphaFold—an AI model capable of predicting complex protein structures based on amino acid sequences. Qiao was dazzled by AlphaFold’s capabilities: there are more possible ways to fold a protein than there are atoms in the universe,

but AlphaFold rapidly predicted protein structures with incredible accuracy. “It was an amazing breakthrough,” Qiao says. “But I could see that two things were still missing from the picture.”

First, AlphaFold2 studies proteins in isolation, but drug discovery requires modeling how proteins interact with smaller molecules. And second, AlphaFold2 offers a static snapshot—but proteins are constantly in motion, bobbing around in the body’s biochemical soup. “You need to capture the full dynamic picture to interpret protein functions,” Qiao explains.

That’s when Qiao had a brainwave: what about using a diffusion model, like those later used in popular GenAI image generators? Much as Michelangelo supposedly carved David by chipping away all the bits of rock that *weren’t* the finished masterpiece, a diffusion model works by gradually eliminating noise that *doesn’t* belong in the target image. Request a picture of a cat in a top hat, and the model takes a clutter of random pixels, then uses its training data—including images of cats and hats—to nudge them closer and closer to the desired outcome.

Applying the same approach to proteins, Qiao realized, would enable modeling of a complex and dynamic molecular landscape. “You start with a random mixture of all possible atomic positions, then progressively de-noise it to reach the final three-dimensional structure,” he explains. Instead of a text prompt, the algorithm is guided by biochemical data: a protein sequence’s evolutionary traits, chemical graphs of small molecules, and a “contact map” predicting how small molecules and proteins should interact. “Using these, the diffusion model maps every single atom to give you the final molecular structure,” Qiao explains.

In layman’s terms, that means Qiao’s prediction engine predicts not just how a protein will fold, but how it will interact with other molecules—including potential pharmaceutical treatments.

Building a computational microscope

Before the rise of AI, modeling a single protein was a monumental task requiring laborious crystallographic or spectroscopic imaging, or equally arduous atom-by-atom computer simulations. “People were using brute force to solve the problem, but modeling took years of GPU time—and still required human labor to set up the simulation,” Qiao explains.

Qiao’s generative models, by contrast, can accurately map complex molecular systems in just seconds. “It’s literally 1 million times faster, and often gives result that are very close to what you’d get from actual experiments,” Qiao says.

In effect, the prediction engine gives researchers a “computational microscope,” letting them peer into the workings of complex biochemical systems with no need for endless calculations or painstaking lab work. It could even help researchers overcome the limits of traditional imaging, and explore the mysterious world of “dark proteins” that haven’t yet been observed experimentally. “By generalizing from known proteins, our model gives us an opportunity to overcome this barrier,” Qiao says. “That could be crucially important as we develop new therapeutics.”

Initial research will likely focus on designing molecules that act on well-understood proteins. “If you already have a biological hypothesis about the signaling pathway you need, you can use our tools to find a molecule to trigger that effect,” Qiao says. “It’s still a non-trivial problem, but our prediction engine accelerates the search process.”

More ambitiously, the engine could be used to design proteins from scratch—to develop novel protein-powered biosensors, for instance, or even to create cascades of bioengineered protein interactions to unlock novel biological functions. One promising potential application would use "molecular glues" to coax disease-causing proteins to bond with other proteins that specialize in degrading biochemicals. "By modeling these complex dynamics, we could actively remove disease-causing proteins from the body," Qiao says. "It's still very early days, but the opportunities are endless."

The ultimate goal is not just to accelerate traditional research, but to automate drug discovery, and propel R&D in directions that human scientists might never have considered. "We hope to build automated workflows that can dream up completely new types of molecules," Qiao says. "In the long run, we'll be able to replace a lot of traditional human-in-the-loop processes in drug discovery."

Beyond academia

To achieve that goal, Qiao has begun looking beyond academia. In 2023, he joined Iambic Therapeutics, a startup spun off from his adviser's Caltech lab; this year, he set out on his own as founding scientist at Chai Discovery. "Academia is a great place to build prototypes—but to have real-world impact, you need a bigger team and more resources," Qiao explains.

The results of this immense teamwork have been impressive: Iambic's NeuralPLexer models set new standards for structure prediction, and Chai's latest model can design novel antibodies with a 16% success rate—a 100X improvement over the previous state of the art. The result, says

Chai Discovery's cofounder Joshua Meier, is a kind of "Photoshop for proteins" that unlocks powerful insights. "Digital biology is no longer science fiction—it's happening now," he says.

Already, DeepMind's AlphaFold database has been used by over 2 million researchers to conduct analyses that would previously have taken 1 billion years to complete; in 2024, two of AlphaFold's creators received the Nobel Prize in Chemistry. Now, Qiao says, researchers are exploring a "post-AlphaFold" world, using more powerful and more specialized models to accelerate discovery. "We have an opportunity to map out protein interactions on an unprecedented scale," he says. "I feel incredibly lucky to have been part of this transformation."

Now, Qiao says, it's time to seize that opportunity, and build the practical tools and workflows needed to deliver game-changing clinical results. "This is an emotional thing for me," Qiao adds. "If we can get this done, then every single thing about the way we do computational drug discovery will be entirely different."