

AI Solves 50-Year-Old Challenge in Biology

Peter Diamandis
Sun, Aug 15, 2021

Thanks to AI, we just got *stunningly* powerful tools to decode life itself.

In two recent back-to-back papers, scientists at DeepMind and the University of Washington described deep learning-based methods to solve protein folding—the last step of executing the programming in our DNA.

Why does this matter?

Because proteins are the building blocks of life. They form our bodies, fuel our metabolism, and are the target of most of today's medicine.

Proteins start out as a simple ribbon of amino acids, translated from DNA, and subsequently folded into intricate three-dimensional architectures. Many protein units then further assemble into massive, moving complexes that change their structure depending on their functional needs at a given time.

And misfolded proteins can be devastating—causing health problems from sickle cell anemia and cancer, to Alzheimer's disease.

In today's blog, we'll discuss the details of this AI-driven advance and what it means for the future of biology and medicine. As our A360 community has often discussed, the biotech field is accelerating, and the decade ahead will bring untold breakthroughs and multi-\$100-billion-dollar startups.

Let's dive in...

(This article originally appeared on SingularityHub by Shelly Fan, adopted by Peter Diamandis for his Abundance Community.)

“A ONCE IN A GENERATION ADVANCE”

One of biology's grandest challenges for the past 50 years has been deciphering how a simple one-dimensional ribbon-like structure turns into 3D shapes, equipped with canyons, ridges, valleys, and caves.

It's as if an alien is reading the coordinates of hundreds of locations on a map of the Grand Canyon on a notebook, and reconstructing it into a 3D hologram of the actual thing—without ever laying eyes on it or knowing what it should look like.

Yes, it's hard. “Lots of people have broken their head on it,” said Dr. John Moulton at the University of Maryland.

It's not just an academic exercise. Solving the human genome paved the way for gene therapy, CAR-T cancer breakthroughs, and the infamous CRISPR gene editing tool.

Deciphering protein folding is bound to illuminate an entire new landscape of biology we haven't been able to study or manipulate. The fast and furious development of Covid-19 vaccines relied on scientists parsing multiple protein targets on the virus, including the spike proteins that vaccines target. Many proteins that lead to cancer have so far been out of the reach of drugs because their structure is hard to pin down.

With these new AI tools, scientists could solve haunting medical mysteries while preparing to tackle those yet unknown. It sets the stage for better understanding our biology, informing new medicines, and even inspiring synthetic biology down the line.

"What the DeepMind team has managed to achieve is fantastic and will change the future of structural biology and protein research," said Dr. Janet Thornton, director emeritus of the European Bioinformatics Institute.

"I never thought I'd see this in my lifetime," added Moulton.

BIRTH OF A PROTEIN

Picture life as a video game. If DNA is the background base code, then proteins are its execution—the actual game that you play. Any bugs in DNA could trigger a crash in the program, but they could also be benign and allow the game to run as usual. In other words, most modern medicine, like gamers, cares only about the final gameplay—the proteins—rather than the source code that leads to it, unless something goes wrong. From diabetes medication to anti-depressants and potentially life-extending senolytics, these drugs all work by grabbing onto proteins rather than DNA.

It's why deciphering protein structure is so important: like a key to a lock, a drug can only dock onto a protein at specific spots. Similarly, proteins often tag-team by binding together into a complex to run your body's functions—say, forming a memory or triggering an immune attack against a virus.

Proteins are made of building blocks called amino acids, which are in turn programmed by DNA. Similar to the Rosetta stone, our cells can easily translate DNA code into protein building blocks inside a clam-shell-like structure, which spits out a string of one-dimensional amino acids. These ribbons are then shuffled through a whole cellular infrastructure that allows the protein to fold into its final structure.

Back in the 1970s, the Nobel Prize winner Dr. Christian Anfinsen famously asserted that the one-dimensional sequence itself can computationally predict a protein's 3D structure. The problem is time and power: like trying to hack a password with hundreds of characters suspended in 3D space, the potential solutions are astronomical. But we now have a tool that beats humans at finding patterns: machine learning.

ENTER AI

In 2020, DeepMind shocked the entire field with its entry into a legacy biennial competition. Dubbed CASP (Critical Assessment of Protein Structure Prediction), the decades-long test uses traditional lab methods for determining protein structure as its baseline to judge prediction algorithms.

The baseline's hard to get. It relies on laborious experimental techniques that can take months or even years. These methods often "freeze" a protein and map its internal structure down to the atomic level using X-rays. Many proteins can't be treated this way without losing their natural structure, but the method is the best we currently have. Predictions are then compared to this gold standard to judge the underlying algorithm.

Last year DeepMind stunned everyone with their AI, blowing other competition out of the water. At the time, they were a tease, revealing little detail about their "incredibly exciting" method that matched experimental results in accuracy. But the 30-minute presentation inspired Dr. Minkyung Baek at the University of Washington to develop her own approach.

Baek used a similar deep learning strategy, outlined in a paper in *Science* this week. The tool, RoseTTAFold, simultaneously considers three levels of patterns. The first looks at the amino acid building blocks of a protein and compares them to all the other sequences in a protein database.

The tool next examines how one protein's amino acids interact with another within the same protein, for example, by examining the distance between two distant building blocks. It's like looking at your hands and feet fully stretched out versus in a backbend, and measuring the distance between those extremities as you "fold" into a yoga pose.

Finally, the third track looks at the 3D coordinates of each atom that makes up a protein building block—kind of like mapping the studs on a Lego block—to compile the final 3D structure. The network then bounces back and forth between these tracks, so that one output can update another track.

The end results came close to those of DeepMind's tool, AlphaFold2, which matched the gold standard of structures obtained from experiments. Although RoseTTAFold wasn't as accurate as AlphaFold2, it seemingly required much less time and energy. For a simple protein, the algorithm was able to solve the structure using a gaming computer in about 10 minutes.

RoseTTAFold was also able to tackle the "protein assemble" problem, in that it could predict the structure of proteins, made up of multiple units, by simply looking at the amino acid sequence alone. For example, they were able to predict how the structure of an immune molecule locks onto its target. Many biological functions rely on these handshakes between proteins. Being able to predict them using an algorithm opens the door to manipulating biological processes—immune system, stroke, cancer, brain function—that we previously couldn't access.

HACKING THE BODY

Since RoseTTAFold's public release in July, it's been downloaded hundreds of times, allowing other researchers to answer their baffling protein sequence questions, potentially saving years of work while collectively improving on the algorithm.

"When there's a breakthrough like this, two years later, everyone is doing it as well if not better than before," said Moulton.

Meanwhile, DeepMind is also releasing (for open and free use) their AlphaFold2 code—the one that inspired Baek.

In a new paper in *Nature*, the DeepMind team described their approach to the 50-year mystery. The crux was to integrate multiple sources of information—the evolution of a protein and its physical and geometric constraints—to build a two-step system that maps out a given protein with stunningly high accuracy.

First presented at the CASP meeting, Dr. Demis Hassabis, founder and CEO of DeepMind, is ready to share the code with the world. "We pledged to share our methods and provide broad, free access to the scientific community. Today we take the first step towards delivering on that commitment by sharing AlphaFold's open-source code and publishing the system's full methodology," he wrote, adding that "we're excited to see what other new avenues of research this will enable for the community."

With the two studies, we're entering a new world of predicting—and subsequently engineering or changing—the building blocks of life. Dr. Andrei Lupas, an evolutionary biologist at the Max Planck Institute for Developmental Biology, and a CASP judge, agrees: "This will change medicine. It will change research," he said. "It will change bioengineering. It will change everything."

FINAL THOUGHTS

This breakthrough demonstrates the impact AI can have on scientific discovery.

And if we couple AI's solution to the protein folding problem with the anticipated breakthroughs in quantum computing—another technology poised to disrupt medicine and healthcare—we're not far from a world where individually customized, precision medicine will move from science fiction to the standard of care.

Nowhere is the convergence of exponential tech bringing greater breakthroughs than in healthcare.

If you're an ambitious entrepreneur interested in healthcare and excited by the progress we've made so far, there's no better time than the present to get involved.

JOIN MY ABUNDANCE360 COMMUNITY

If you want my coaching as an exponential entrepreneur and to reinforce those mindsets that will inspire and guide you to create a hopeful, compelling and abundant future for yourself and humanity, then consider joining my year-round Abundance360 coaching program.

Every year, my team and I select a group of 360 entrepreneurs and CEOs to coach over the course of a year-long program. You can join at any time for a full year's worth of personal coaching on Exponential Tech, Longevity, Moonshots, and Abundance thinking.

My mission is to help A360 members **obtain mastery in four specific mindsets:** an Abundance Mindset; an Exponential Mindset; a Moonshot Mindset; and a Longevity Mindset. Together we will actively select and reinforce your preferred Mindsets.

To learn more and apply to A360, visit abundance360.com