2 AlphaFold

Prediction without Understanding

It's an unfortunate reality, for those of us who write about AI, that the 14th biennial Critical Assessment of Structure Prediction (CASP) meeting took place over Zoom. So many other epochal events in the history of science can be evoked with vivid set pieces – Guglielmo Marconi braving Newfoundland's tempests to receive the first transatlantic radio signal; Marie Skłodowska Curie admiring the unearthly glow of radium in the shed that served as her laboratory; Newton witnessing the fall of an apple from a tree. The image of scientists crouched over their laptops in home offices scattered around the world, during the eighth month of a global pandemic, doesn't quite measure up.

But what those scientists saw on their computer screens on November 30, 2020 eminently merits a spot in the trophy case of human scientific achievement, for all that it lacks in romance. At CASP 14, deep neural networks became full members of scientific society.

CASP, which was founded in 1994 by the protein scientists John Moult and Krzysztof Fidelis, is both a conference and a competition, and its goal is to evaluate approaches for predicting the three-dimensional structures of proteins from their amino acid sequences. Proteins are microscopic molecules that carry out or contribute to almost every biological

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

process: They shuttle materials around the cell, assemble and break down other molecules, and copy our DNA so that our cells can divide, among numerous other responsibilities. Each of these jobs is undertaken by proteins with specific, specialized structures, so knowing those structures can help us figure out how those processes work and determine why they sometimes go wrong. And when proteins do harm, ascertaining their structures can be key to stopping them: In the 1990s, scientists used their knowledge of the structure of a key HIV protein called protease to design drugs that stopped it from functioning, and those drugs helped transform HIV from a death sentence into a chronic but manageable condition.

For decades scientists have determined protein structures experimentally, first with X-rays and more recently with electron microscopes. But these approaches are arduous and time-consuming – it can take months, or even years, to determine the structure of a single protein. At the time of the CASP 14 meeting, scientists had experimentally solved the structures of only about 100,000 proteins, a tiny fraction of the billions whose amino acid sequences were known. A reliable computational method to predict the structure of a protein, and avoid all of the hard work and waiting involved in experimental structure determination, would revolutionize biology.

On the first morning of CASP 14, Moult bucked tradition and spoiled the competition's results because the news was too big to wait: After twenty-six years, a reliable structure prediction method had at last been found. For protein structure researchers, this was a field-defining moment, something many had spent their careers waiting

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

for. "I always hoped I would live to see this day," Moult later told the *New York Times*. "But it wasn't always obvious I was going to make it."[1] The eminent bioinformatician Janet Thornton shared a similar sentiment with the *Guardian*: "This is a problem that I was beginning to think would not get solved in my lifetime."[2]

What made this moment so notable was not only the fact that a solution now existed – an enormous accomplishment in its own right – but also the specifics of that solution. The winning method, called AlphaFold, was a deep neural network-based system designed and trained by the Google subsidiary DeepMind, which had made international headlines back in 2016 when one of its deep neural network systems beat Go world champion Lee Sedol. Computer scientists had long held up the deceptively complicated board game of Go as a major challenge for artificial intelligence, so conquering it was a feather in DeepMind's cap.

But winning board games doesn't save lives. With AlphaFold, DeepMind, and the deep learning community as a whole, made a powerful new argument for its existence: If deep neural networks could conquer major scientific challenges then perhaps they might one day design drugs for untreatable illnesses, develop greener energy technologies, or predict natural disasters long in advance. AlphaFold was AI at its most virtuous.

In the years since CASP 14, AlphaFold has remained the preeminent example of deep learning in science, even as large language models have come to dominate much of the rest of the AI conversation. It is regarded so highly in the scientific community that it earned John Jumper, who led

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

the project, and Demis Hassabis, DeepMind's CEO, the 2024 Nobel Prize in Chemistry. That makes AlphaFold a unique project, but also an illustrative one. It represents the greatest hopes of what AI might do for science, and for the world.

Through understanding AlphaFold – how it works, what it does, and what it cannot do – we can start to grasp the promise, and the pitfalls, of deep neural networks in science. In particular, AlphaFold is a stark demonstration that the scientific understanding of a process is often completely irrelevant to a deep neural network's ability to make predictions about that process. At the time when AlphaFold was built, scientists did have a solid theory about why proteins fold into particular shapes, but the role this understanding plays in AlphaFold's predictive success is limited. And so, just as AlphaFold's tectonic impact on the fields of molecular and cellular biology indicates how valuable a fast and reliable predictive system can be, the specifics of its design make clear that understanding may now be superfluous for such systems. AlphaFold's creation may very well mark the first time that scientists have managed to so thoroughly dissociate prediction and understanding.

Steps Toward Understanding

In order to grasp what AlphaFold means for the study of proteins, and for science as a whole, it's important to have a sense of the intellectual terrain into which it emerged. By CASP 14, protein folding was a mature subdiscipline of chemistry and biophysics, and some scientists thought that the field's biggest questions had been answered. One of those

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

scientists is Peter Wolynes, a chemist at Rice University who has made major contributions to our contemporary understanding of protein folding. "I viewed the problem of protein folding as being intellectually solved," he says, "somewhere around the beginning of the 2000s."

That progress happened over a comparatively short span of time: Unlike the solar system, which the Greco-Roman astronomer Ptolemy strove to understand almost two millenia ago, proteins are a modern topic. The word "protein" wasn't introduced until 1838, and then only to describe an aggregate substance present in blood, eggs, and plant matter, rather than the individual molecules that constitute it.

Much of our current knowledge of proteins can be traced back to the German chemist Emil Fischer, who first aimed his attention toward them around the turn of the twentieth century. At that time, scientists knew that proteins could be broken down into much smaller subcomponents called amino acids, but they were unsure precisely how amino acids combined to form proteins. Fischer theorized that amino acids can link up one after another like pearls in a strand, and he proved it by synthesizing short amino acid chains himself.

Today, we know that human proteins are made up of twenty distinct amino acids, and scientists have assigned each of these amino acids a different letter of the alphabet. The specific sequence of these twenty amino acids that a protein comprises is known as its primary structure, and scientists record a protein's primary structure by writing down the corresponding sequence of letters.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

Though proteins are, at their core, just flexible chains of amino acids, most of them don't loosely float about the cell, undulating like an inflatable tube man outside a gas station. Once synthesized, most amino acid chains twist and curl on themselves until they achieve a stable, three-dimensional structure, and it is this structure that determines a protein's job in the cell. For example, enzymes – globular proteins that speed up chemical reactions – have regions on their surfaces known as active sites, which are the right shape to cradle specific molecules. Each enzyme will gladly break apart or mold together the molecules that it evolved to interact with but will have limited effects on other molecules.

In the case of enzymes, the structure–function relationship is fairly straightforward. But our cells are also home to huge protein complexes that function so intricately they've earned the name "molecular machines." Take ATP synthase, the protein complex that converts energy obtained from metabolized food into a storage molecule called ATP. If you could shrink down to the size of a virus and watch ATP synthase at work, you'd see something that looks a bit like an old-fashioned butter churn, with a small water wheel stuck on the end of its handle. The water wheel is embedded in a membrane; when positively charged atoms flow across that membrane, the water wheel turns, and the handle rotates. Meanwhile, the body of the churn remains still, and, like butter, ATP is created where the moving handle meets the stationary body.

There's also kinesin, a protein with two foot-like protrusions that walks along long filaments to transport substances throughout the cell, and DNA helicase, a

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

donut-shaped protein that encircles a single strand of DNA to unzip its double helix. In each case, the three-dimensional structures of these protein complexes, and thus the structures of the protein subunits that make them up, are essential to their ability to do useful work in the cell.

But in the early days of protein science, no one had the slightest idea what any specific protein looked like. Knowledge of proteins on the whole was so uncertain that it took decades for anyone to fully connect Fischer's discovery of amino acid chains to the question of protein shape. That essential step forward was made by Linus Pauling, an American chemist who would later become just as renowned for his nuclear disarmament advocacy as for his contributions to science. He is the only person ever to receive two unshared Nobel prizes – the Nobel Prize in Chemistry and the Nobel Peace Prize.

In 1936, Pauling and his colleague Alfred Mirsky published a paper arguing that the three-dimensional structure of a protein is determined by chemical interactions among different portions of its amino acid chain.[3] This went well beyond Fischer's discovery that each amino acid in a chain is strongly bonded to the acids before and after it: Pauling was proposing that a variety of weaker connections among more distant amino acids could fold the chain up into a specific structure and keep it held in place. According to Pauling, the locations in which these bonds formed, and therefore the three-dimensional structure of the protein, came down to the chemical structure of each amino acid.

A couple of decades later, the American chemist Christian Anfinsen strengthened Pauling's theory through

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

a series of experiments that he performed on the enzyme ribonuclease, which he obtained in large quantities from a Chicago meat packing company that extracted it from cow pancreases.[4] First, Anfinsen added bond-breaking chemicals to test tubes of purified ribonuclease. That rendered the ribonuclease unable to perform its job – breaking down the molecule RNA – so Anfinsen knew that its three-dimensional structure had been disrupted. Next, he removed those chemicals and exposed the unraveled proteins to air, thereby allowing their bonds to reform. Finally, he tested whether the reformed proteins could once again break down RNA, and found that they could. Anfinsen concluded that the proteins had snapped back into their initial structure without assistance from the cellular machinery that would typically surround them. Somehow, a protein's three-dimensional structure was encoded in its amino acid sequence.

Together, Pauling's theories and Anfinsen's experiments made articulable the problem that, a century later, AlphaFold would solve: Can you translate the string of letters that describes a protein's amino acid sequence into the three-dimensional structure that it will adopt? Neither Pauling nor Anfinsen made much progress on this prediction problem, and by the time they both died, in the 1990s, it was still decades from being solved.

But they did both make major contributions to the why question of protein folding: Why do many amino acid chains fold into specific structures rather than the near-infinite alternative structures that they might conceivably take on? For years, Pauling applied his deep understanding of chemical bonds to this question while also grappling

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

with the political fallout of his advocacy work, which ultimately led to him being labeled a Communist. He eventually determined that many proteins contain corkscrew-like substructures known as alpha helices and accordion-like substructures known as beta pleated sheets, and that both are held together by a type of chemical bond known as a hydrogen bond. Together with his colleagues, Pauling published these discoveries in 1951.[5] From that point on, scientists could reasonably say that they understood some aspects of protein folding – they had an answer to the question of why amino acid chains form alpha helices and beta sheets.

In the years that followed, scientists continued to chisel away at the problem of protein structure. Most notably, Walter Kauzmann, another American chemist, proposed in 1959 that amino acids with positively or negatively charged regions will sit at the outside of the protein, whereas amino acids without charged regions try to stay sequestered toward the center of the molecule.[6] His explanation was that amino acids with charged regions can form hydrogen bonds with surrounding water molecules, but those without charged regions prefer, like oil, to stay far away from water.

It was Anfinsen who eventually assimilated Pauling's research on hydrogen bonds and newer ideas about charged and uncharged regions, as well as his own work with ribonuclease, into a coherent theory of why proteins adopt particular structures. That theory, which Anfinsen called his "thermodynamic hypothesis" of protein folding, is so influential that it is now often known as Anfinsen's dogma.[7]

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

The thermodynamic hypothesis hinges on a quantity known as free energy. Free energy is a property of some physical systems, including proteins; and as those systems change and evolve, their free energies may change as well. It's often helpful to use height or elevation as a metaphor for free energy, and indeed, scientists do so all the time. That's because, in the same way a ball placed up high on a slide will effortlessly roll down to the bottom, a physical system will automatically transition into a lower free energy state if it has the opportunity to do so.

In the case of amino acid chains, free energy is something like a measure of chemical discomfort. Some chemical interactions, such as the interactions between uncharged amino acids and water molecules, are uncomfortable and increase a protein's free energy. Others, like hydrogen bonds between amino acids, decrease it. Like a bored student sitting in an uncomfortable chair, chemical systems will shift and evolve naturally to minimize the discomfort of high free energy. That's essentially what Anfinsen's thermodynamic hypothesis proposes: Proteins fold into shapes that minimize their free energy, and if there's only one energy-minimizing shape then that's the shape that the protein will always adopt.

The thermodynamic hypothesis provided the first general answer to the question of why amino acid chains fold into specific shapes. Whereas Pauling understood some specific aspects of protein folding, Anfinsen understood the phenomenon as a whole, albeit at a very high level. But the hypothesis still leaves a lot to be desired. For one thing, it's not detailed enough to help scientists predict the shapes that

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

amino acid chains will fold into. For another, it has nothing at all to say about *how* an amino acid chain finds that ideal, minimum free energy shape.

While Anfinsen was publishing his studies on ribonuclease and advocating for his thermodynamic hypothesis in the 1960s, the American molecular biologist Cyrus Levinthal recognized that the how question was a major problem for that hypothesis.[8] In order for a protein to be sure it had found its minimum energy shape, he reasoned, it would have to try out all of the shapes it could possibly fold into. But according to his estimates, a relatively small protein might be able to adopt as many as 10^{300} different shapes, and trying out all of those shapes would take longer than the age of the universe. Anfinsen's dogma now had a counterpart in Levinthal's paradox.

The Folding Funnel Theory

That paradox was still defining the mood in the late 1980s. That's when Wolynes, who was a decade out from a PhD in chemical physics and working as a professor at the University of Illinois, decided to turn his attention toward protein folding. To his peers, it probably looked like a foolhardy career move. "People were saying at the time, 'Protein folding is an insolvable problem,'" he says.

But Wolynes had come across an idea in a different area of chemistry that he thought might prove fruitful: the free energy landscape. Energy landscapes take the analogy between free energy and height to its utmost extreme. When illustrated in scientific papers and textbooks, they look a lot

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

like physical landscapes dotted with hills and valleys – but instead of representing a location on a map, each coordinate in an energy landscape is associated with a different state of a physical system. In the case of proteins, every spot on the free energy landscape corresponds to a specific shape that the amino acid chain could theoretically fold into, with similar shapes located close to each other. The height of the landscape at each spot reflects the free energy of that shape, so the highest-energy, most uncomfortable shapes will appear as mountains on the landscape, and the settled, low-energy shapes will look like valleys. The minimal-free-energy shape – which, according to Anfinsen, is the shape that the protein will ultimately adopt – will be the deepest valley of all.

The free energy landscape idea is useful because it offers an intuitive way to grasp not only why but also how proteins fold. According to this picture, an amino acid chain is like a ball that has been dropped on an energy landscape. In the same way the ball would roll downhill over the landscape until coming to rest in a valley, the chain will twist and flex into lower and lower energy shapes until it reaches its final folded state. Because of the inherent randomness of the protein folding process, the analogy isn't perfect: The protein chain will sometimes transition from lower- to higher-energy shapes on the way to its final state. So a more precise comparison might be a ball rolling across a landscape during a mild earthquake; the shaking might be enough to pop the ball out of a small surface depression, but it will still ultimately settle somewhere, as long as that place is deep enough.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

When Levinthal had proposed that a protein might have to fold into all its possible confirmations in order to find the one with the lowest energy, he had implicitly been imagining an energy landscape that resembled a golf course – relatively level, with a deep hole at the location corresponding to the folded state. In order to find that one dramatically lower-energy shape, the protein would have to wander randomly around the landscape for eons.

Wolynes had a different idea: What if the free energy landscape guided the protein to where it needed to go? Rather than picturing a flat landscape with one hole, he envisioned a landscape that looked more like a big funnel leading down to the folded state. As the amino acid chain folded, it would essentially roll down the funnel until it reached that energy minimum. Together with a graduate student, Wolynes published a paper detailing the mathematical underpinning of this theory in 1987; that same year, Ken Dill, then a professor of pharmaceutical chemistry at the University of California San Francisco, wrote that protein folding was more a problem of "funnels than tunnels" in a book chapter.

In one fell swoop, the folding funnel theory resolves Levinthal's paradox and validates Anfinsen's dogma. Simultaneously, it raises a glaring question: Where do those funnel-shaped energy landscapes come from? There's no fundamental law of physics or chemistry that suggests that amino acid chains should all have funnel-shaped energy landscapes. The answer, instead, lies in biology. A random chain of amino acids will almost certainly have a far more rugged and disordered energy landscape than that which Wolynes envisioned, but our cells do not contain random chains of

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

amino acids – every protein in every organism is the product of billions of years of evolution. According to the folding funnel theory, only proteins with a particular kind of energy landscape will fold quickly and reliably into a stable three-dimensional structure. Amino acid chains without such energy landscapes will tend to be removed from the gene pool.

Over the past few decades, the folding funnel theory has gone mainstream, and illustrations of funnel-shaped energy landscapes appear in countless scientific articles on protein folding. Those illustrations furnish real understanding. Armed with the intuitive, visual metaphor of the funnel, even protein folding neophytes can answer questions about why and how proteins fold without too much difficulty. And those who grasp the underlying math and chemistry in more depth can make detailed predictions about how changing a protein's energy landscape might affect the rate at which it folds – or whether it successfully folds at all – under various conditions. For many scientists, including Wolynes, that's enough to say that we understand protein folding.

But, much like Kepler's laws, the folding funnel theory leaves many questions unanswered. "A complete understanding of protein folding should lead to a scheme for predicting three-dimensional protein structure from one-dimensional sequence information," wrote Wolynes in his 1987 paper. By those lights, the understanding supplied by folding funnel theory is incomplete. The theory does tell us that we can find a protein's folded state at the bottom of a funnel-shaped valley in its free energy landscape, but that information is only useful if you can reliably construct the energy landscape for a given amino acid chain.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

So that's what scientists tried to do. Through the 2000s and the 2010s, top-tier structure prediction algorithms combined physics-informed free energy calculations with clever computational tricks. For a while, it seemed like figuring out how to translate amino acid chemistry into energy landscapes might be a key step on the way to achieving protein structure prediction. As has so often been the case in science, prediction would follow from understanding.

The Structure Prediction Problem

It was arguably Pauling who pioneered the discipline of protein structure prediction. When he published his studies on alpha helices and beta pleated sheets in 1951, no protein structures had yet been solved – the alpha helix and beta sheet were predictions about what scientists might one day observe, once their experimental techniques had improved. Their accuracy is a testament to just how well Pauling understood chemical bonding.

But though our chemical understanding has advanced considerably since Pauling's day, it alone still isn't enough to predict something as complex as the full structure of a protein. For decades, successfully predicting protein structures required weaving together chemical understanding and known protein structures in increasingly clever ways – until, finally, AlphaFold left the chemical understanding piece behind.

The first attempts to predict the structures of complete proteins – rather than the structures of parts of proteins, as Pauling had done – leaned heavily on protein

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

structure data. In 1969, researchers realized that a protein called α-lactalbumin, whose structure was unknown, shares 40 percent of its amino acid sequence with a protein called lysozyme, whose structure had previously been solved using a technique called X-ray crystallography.[9] They already had a brass model of lysozyme, measuring roughly a meter across, which they had built out of Tinkertoy-like pieces that were mass manufactured for chemists. In order to transform the model of lysozyme into a model of α-lactalbumin, they identified regions where the two proteins' sequences differed substantially and then used their knowledge of amino acid bonding to guess at how those regions might fold in α-lactalbumin. Later work confirmed that the hypothetical α-lactalbumin model they ended up with was impressively accurate.

Several years later, a resource called the Protein Data Bank (PDB) was established in order to make valuable protein structure data more accessible to scientists. When it began formal operations in 1973, it stored the structures of a mere nine proteins and would distribute data to interested scientists by shipping out reels of magnetic tape. Today it archives over 250,000 protein structures, and you can download any of them in a matter of seconds.

Making progress on protein structure prediction would in all probability have been impossible without a resource like the PDB. For one thing, PDB structures offer an essential way to test new prediction approaches: The best indication of whether a structure prediction approach will work well in the wild is its ability to predict structures that have already been experimentally confirmed. When the

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

CASP competition was founded, it formalized this use of known protein structures.

For another, the PDB provides scientists with the large-scale protein structure data that powers all modern protein structure prediction approaches. In the pre-AlphaFold days, the tactic first used to solve the structure of α-lactalbumin was still alive and well under the name of homology modeling, and the volume of known protein structures in the PDB had made it far more powerful than it was back in the 1960s. Scientists would compare the sequence of an unknown protein against the sequences in the PDB, and if they found a good match – which became ever likelier as more sequences were added to the PDB – they assumed that their protein's structure was identical to the PDB protein's structure everywhere that the sequences corresponded. Understanding played an essential role in homology modeling as well: In order to fill in the sections of the protein where the two sequences didn't line up, scientists could use software to approximate the free energies associated with the various possible shapes that those regions could take on.

Homology modeling works well when substantial portions of a target protein's structure match a known structure. But without a good match, it falls apart, because it is extremely difficult to determine large swaths of a protein's structure solely on the basis of our current understanding of physics and chemistry. The math involved is just too difficult and time-consuming, even for a powerful computer. In order to fill that gap, the structure prediction community focused much of its effort on devising strategies for predicting the

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

structures of proteins that don't have solid PDB matches. Like homology modeling, these approaches drew heavily on both protein structure data and the scientific understanding of protein folding.

Before AlphaFold, one of the most successful methods to determine the structure of proteins without good PDB matches was a piece of software called Rosetta. Pioneered by University of Washington biochemist David Baker, Rosetta takes the core insight of homology modeling and scales it way, way down. Instead of trying to match the entire target amino acid sequence to a PDB sequence, it breaks down the target sequences into short chunks of three or nine amino acids and then searches for places where those short sequences appear in known proteins. That gives the software a menu of structures that each of those short sequences might conceivably fold into. Then, the software attempts to assemble these fragmentary structure puzzle pieces into full protein structures in a variety of different ways.

That's how Rosetta uses data. But understanding is equally, if not more, essential to its success. In order to calculate the energy of candidate protein structures, Rosetta uses a complex function that takes into account hydrogen bonding, the tendency of some amino acids to stay sequestered from water, and various other chemical phenomena that contribute to the free energy of folded proteins. When Rosetta is assembling puzzle pieces into full structures, it uses that energy function to select which pieces it should add next. It runs this process many times to generate multiple candidate structures for each amino acid sequence, and

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

then it chooses from among those structures by calculating the energy of each one.

In this final step, Rosetta draws on folding funnel theory. It searches for clusters of similar candidate structures that all have low energies, under the assumption that such a cluster indicates the presence of a large valley – a funnel – in the free energy landscape. The minimum free energy structure can then be found at the bottom of the funnel.

Baker and his lab used Rosetta to score near the top of the pack at CASP several times. Still, before the 2020s, none of the available approaches was anywhere near solving the structure prediction problem, especially for proteins without good PDB matches. A biologist who needed to know the structure of a new protein for their research couldn't depend on Rosetta – they'd have to send the protein to an electron microscopist or X-ray crystallographer and wait months for an answer.

When the 13th edition of CASP rolled around in 2018, Baker had been incrementally improving Rosetta for almost two decades. By that point, he says, he was ready to see something new. And, that year, he did. The top-scoring approach at CASP 13 was an early version of AlphaFold.

The AlphaFold that DeepMind put forward that year didn't decimate the competition. But its first-place finish did attract some press coverage, and it stunned the academic structure prediction community. In a blog post titled "AlphaFold @ CASP13: 'What just happened?'", the computational biologist Mohammed AlQuraishi admitted to feeling some amount of melancholy in the face of AlphaFold's success. "In a delicious twist of irony, we the people who have

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

bet their careers on trying to obsolete crystallographers are now worried about getting obsoleted ourselves," he wrote.[10]

Baker, however, was energized: He and his lab members had already been dipping their toes into deep learning, and AlphaFold's success inspired them to start working on it in earnest. By the time DeepMind won CASP 14 with a revamped version of AlphaFold, the group was ready to apply what they had learned at the meeting to build their own protein folding neural network. This was a completely new approach to protein structure prediction – so, despite the decades of effort and wealth of scientific understanding that had gone into Rosetta, Baker and his team put their old tool entirely to the side and built its successor from scratch. In mid-2021, the Baker lab released RoseTTAFold, its newest and best piece of protein folding software; at almost exactly the same time, DeepMind made the code for AlphaFold publicly available.

What made AlphaFold, and RoseTTAFold, so much more successful than the protein structure prediction algorithms that had come before? The answer, in short, is data, plus the means to use it effectively. By the time AlphaFold was developed, the PDB had archived over 100,000 protein structures. While traditional structure prediction tools such as Rosetta did draw on PDB data to make their predictions, they were still leaving tons of useful information on the table. By instead using a deep neural network, DeepMind got much more mileage out of all the PDB data.

But not just any deep neural network would do. In order to make the best possible use of the available protein structure data, the DeepMind team designed AlphaFold

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

to incorporate a variety of clever ideas about the structure prediction problem. For instance, the deep neural network at the center of AlphaFold doesn't just receive an animo acid sequence and spit out a structure. Before running the network, AlphaFold follows the homology modeling handbook and compares the target sequence against PDB sequences; the geometric details of the structures of the few closest PDB matches are then fed into the neural network. Simultaneously, AlphaFold searches a much bigger database that contains amino acid sequences, but not protein structures, from all sorts of different species. The goal of that step is to use information about how a sequence evolved to make inferences about its structure: If two amino acids change in lockstep throughout a protein's evolutionary history then perhaps they are located close together in the folded protein and need to be able to fit together cleanly. The results of this evolutionary search are also fed into the deep neural network.

What AlphaFold doesn't take into account is the hard-won physical understanding of protein folding achieved by scientists such as Anfinsen and Wolynes. The system's designers never explicitly used any information about free energy in the design or training of AlphaFold's core network – probably because they assumed, or knew, that such information would not improve its performance. In its final step, AlphaFold does briefly simulate the forces among the various molecules in the protein. But this step does nothing to improve the accuracy of the predicted structure, when measured objectively – it just makes the structure look better to human eyes.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

For decades, scientific ideas about how and why proteins fold were an essential element of the protein structure prediction process. AlphaFold left those ideas behind. That's not to suggest that AlphaFold is unscientific: The way that it incorporates protein sequence data from various species, for example, only makes sense if one understands the basics of evolution. But it is notable that AlphaFold was designed to predict the outcomes of a physical process that is reasonably well understood, and yet it makes essentially no use of that understanding. If Anfinsen had never dreamt up the thermodynamic hypothesis, and if Wolynes and Dill and their peers had never started talking about folding funnels, AlphaFold might still be just as successful as it is today.

The Fruits of Structure Prediction

In the words of John Jumper, AlphaFold sparked "an era of structural abundance" in molecular biology. Once he and his team released the code for AlphaFold in 2021 – and especially once they released a database of 200 million predicted protein structures in 2022 – AlphaFold quickly and thoroughly embedded itself in the pipelines of laboratories worldwide. The paper that describes AlphaFold's architecture has now been cited about forty thousand times, and over three million people have used the AlphaFold protein structure database.[11]

Every scientist I spoke with while researching this chapter emphasized the immense and invaluable contribution that AlphaFold has made to molecular biology, and

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

most of them said that they regularly use AlphaFold in their own work. In the pre-AlphaFold days, to learn what a protein looked like, they would have had to wait for laborious, time-consuming structure determination experiments to be completed. Now, they can obtain in minutes structures that would previously have taken months.

This transition from relative structural scarcity to mind-boggling structural wealth is entirely attributable to deep neural networks, and it may represent the most visible impact that neural networks have had on science so far. But the structures themselves aren't the point. Dozens of researchers from around the world devoted decades of their lives to CASP not because it's exciting to see the structures of the tiny molecules that keep us alive – though it certainly is – but because protein structures can be put to numerous scientific purposes, both practical and intellectual. And examining the purposes to which scientists have in fact put AlphaFold in the past few years provides hints about the effects that neural networks might have in other scientific fields.

On their own, the predictions that AlphaFold makes don't have the immediate, real-world applicability of, say, a medical diagnosis prediction. But they open up promising avenues, particularly in drug development. Most drugs work by bonding with proteins in order to interfere with, or sometimes enhance, their functioning – statins, for instance, physically block an enzyme that helps the cell manufacture cholesterol. When drug makers want to identify molecules that will interact with a protein that plays a role in some disease, they often use computer methods to

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

search for molecules that will fit with the protein's structure in the right way. Before AlphaFold, this approach required an experimentally determined structure, but recent research has shown that AlphaFold structures can be just as useful for identifying drug candidates.

But AlphaFold's greatest practical worth may lie in its ability to make counterfactual structure predictions. Just as Newton's laws allow physicists to reason about how planets would move in solar systems nothing like our own, AlphaFold can guess the structures of protein that don't actually exist. That means it's possible to use AlphaFold and related tools to design completely novel proteins.

Protein design isn't a new field: David Baker was awarded the 2024 Nobel Prize in Chemistry, alongside Hassabis and Jumper, for his work in protein design, much of which was completed over a decade before AlphaFold won CASP. Back in 2003, Baker and his team used Rosetta to design an amino acid chain that folded into a shape that had never been observed in nature – the first time anyone had designed such a protein.[12] They've since created enzymes that accelerate specific chemical reactions, proteins that bind to specific small molecules, and proteins that assemble themselves into huge, complex structures – all of which have various potential medical applications. They even designed proteins that, when administered to mice in a nasal spray, helped prevent and treat SARS-CoV-2 infection.

So when AlphaFold, a far better structure prediction tool than Rosetta, emerged on the scene, Baker got to work adapting its approach for protein design. By retraining RoseTTAFold to predict protein structures not

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

from specific amino acid sequences but from general constraints – a molecule that the protein should be able to bind with, for example, or the particular kind of symmetry it should have – Baker's team developed a protein design neural network that they called RFdiffusion. When tasked with designing new proteins that could bind to proteins with known structures, RFdiffusion blew the team's earlier, physics-based protein design strategy out of the water. The deep learning-based approach is "just qualitatively better," Baker says.

In the past two years, Baker's lab has used RFdiffusion to design antivenoms for deadly snakebites, molecules that assemble into fibers whenever they're exposed to a high-enough pH, and numerous other proteins. And Baker is far from the only person applying deep neural networks to protein and drug design. Soon after AlphaFold's CASP 14 victory, Hassabis founded Isomorphic Labs, a sister company of Google DeepMind that is explicitly committed to discovering new drugs using AlphaFold. As of early 2026, Isomorphic is reportedly gearing up for human trials of some AI-designed drugs – though it is notable that, so far, no drugs designed with AlphaFold have made it to market. But when a drug designed using AlphaFold or RFdiffusion eventually does become available, deep neural networks will have proven their practical worth beyond a shadow of a doubt.

AlphaFold is thus poised to become an unmitigated victory for those who advocate for the practical view of science – and anyone who ever needs to take a drug. But even a stolidly intellectual, incorporeal being could find aspects of

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

AlphaFold to like. The tight link between protein structure and function means that figuring out a protein's structure can often unlock new insights about how it performs its job within the cell.

Here's an example of what that looks like. In 2022, only a year after Google released its first batch of AlphaFold-predicted structures, a team of researchers from the Netherlands Cancer Institute published a paper on the function of a previously unstudied protein that they called ACTMAP.[13] The team knew that ACTMAP likely assisted in the formation of actin, a key protein in muscle tissue, but they weren't yet certain how. Though ACTMAP's structure had never been experimentally determined, they found a predicted structure in the AlphaFold database – and, based on that structure, they were able to infer that ACTMAP's job is to snip off an extra amino acid from immature actin molecules, a hypothesis that they subsequently confirmed with experiments.

There's no question that this study, and others like it, represent small but meaningful steps forward for scientific understanding. Thanks to AlphaFold, scientists now know more about how actin forms than they used to. That's because protein structure predictions are not merely descriptions of what biological objects look like; They are also stories about how and why they function. And while ACTMAP's role in actin production might have been sussed out in the pre-AlphaFold days if the team had asked their friendly neighborhood X-ray crystallographer to determine its structure for them, doing so would have added months to their research timeline. AlphaFold hasn't single-handedly

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

enabled us to understand how and why the minuscule machines inside our cells do what they do, but it has made the process of acquiring that understanding much faster.

But AlphaFold isn't a universal accelerant for understanding. Even though it enables scientists to more easily answer how and why questions that depend on protein structures, so far it has been of limited use for providing insight into structures themselves. There might be some reason to think that studying how AlphaFold works could one day advance our knowledge of protein folding: Some researchers have argued that AlphaFold makes its predictions by minimizing something that resembles an energy function.[14] Either way, we don't yet have a sturdy enough sense of how deep neural networks work to glean much from looking at AlphaFold.

For his part, Jumper is skeptical that AlphaFold will ever tell us much about the how and the why of protein folding. "I don't expect we will rewrite the protein science textbooks based on what we find there," he says. AlphaFold might be like the oracle who makes her predictions using a barometer: It may have uncovered an effective predictive strategy, but that strategy could have as little to do with the protein folding process as reading numbers off a barometer has to do with climate physics.

Jumper, however, isn't too concerned about AlphaFold's likely inability to help us with these questions, because he has a long history with unilluminating computational tools. Before he turned his attention to deep learning and joined DeepMind, Jumper worked on molecular dynamics software that aimed to predict how an animo acid chain

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

would fold by simulating every one of its atoms. The molecular dynamics tools he used were much less versatile, and far more computationally intensive, than AlphaFold – and he didn't understand how they worked, either. Sure, he knew the formula to calculate the forces on a given atom, and he knew how the atom would respond to those forces, but he can also calculate how individual neurons in the AlphaFold network will behave if he knows their inputs. None of that knowledge helps him understand how and why proteins fold. From that perspective, AlphaFold's distant relationship with understanding is nothing new.

But there are a couple of key differences between AlphaFold and molecular dynamics simulations. For one, it seems likely that a good molecular dynamics simulation could further scientific understanding, even if it's difficult to grasp how it works in its entirety. Molecular dynamics simulations aim to track how atoms behave in the real world with exacting precision, so a reliable simulation would allow scientists to study the protein folding process in far more detail than has ever been possible using experimental techniques. It's not tough to imagine that a wealth of high-quality, simulated data about the protein folding process might advance our theoretical understanding of that process.

The second, and more important, difference is that molecular dynamics simulations didn't solve the protein structure prediction problem, and AlphaFold did. Deep neural networks are not the first, or the second, or likely even the hundredth computational method to achieve predictive success and simultaneously defy human understanding. What distinguishes them is the magnitude of their

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

successes. That makes them transformative in a way that previous technologies were not.

In Defense of Protein Folding

After AlphaFold solved the protein structure prediction problem, news of the achievement spread fast. But not all of it was accurate. Following the lead of DeepMind's blog, several news outlets, including the *New York Times* and the *Daily Mail*, proclaimed that AlphaFold had "solved the protein folding problem."[15]

It's widely accepted in the protein folding field that the "protein folding problem" is actually multiple separate problems. In an influential 2008 paper, Ken Dill and his colleagues articulated those as three problems: How does a sequence of amino acids encode a particular three-dimensional structure? What is the pathway, or the series of steps, by which an amino acid chain folds up into its final form? And, is it possible to predict a protein's structure solely based off of its amino acid sequence?[16]

That third question has now been answered with a resounding yes. But, as Dill and his colleagues made clear in their paper, those questions are not dominoes that fall down one after another. Declaring the protein folding problem completely solved upon the development of a viable structure prediction method would be like Goldilocks abandoning the bears' house after discovering that one bowl of porridge was too hot.

The idea that AlphaFold solved the protein folding problem writ large, rather than the protein structure

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

prediction problem, is now so widespread that researchers who work on other aspects of protein folding have had to explicitly defend their work against that misapprehension. Greg Bowman, a biophysicist at the University of Pennsylvania who has worked on protein folding for the past two decades, has always been more interested in the how and why of protein folding than in the structure prediction problem, so he never paid much attention to CASP. But he realized he needed to be looking into AlphaFold when people started asking him whether it would put him out of a job. Bowman clarified his thoughts on the matter with the title of his next paper: "AlphaFold and Protein Folding: Not Dead Yet!"[17]

Other protein folding scientists are also protecting their turf: The Johns Hopkins biophysicist Stephen Fried has a section on the front page of his lab's website titled "Hasn't AlphaFold Already Solved This?" (The answer? "We don't think so!") "We have had to maybe defend our existence a little bit more," Fried says.

In fairness, the idea that AlphaFold solved protein folding is more than just a mistake or a canny bit of self-promotion from Google. AlphaFold really did solve the protein folding problem for some scientists. Even the biophysicist George Rose, as native a denizen of the protein folding world as one could hope to find – he published his first papers on protein folding only a few years after Anfinsen formally articulated his thermodynamic hypothesis – understands why some people might feel that the protein folding problem is now a thing of the past. To a molecular biologist who wants to piece together how a protein functions

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

inside the cell, he says, the how and why of protein folding don't matter much; the problem is solved for them once they receive an accurate structure prediction.

But Rose sees things differently. "Solved for me means that I understand the physical chemistry underpinnings of how [the folded protein] got that way," he says.

None of the protein folding experts I spoke with believe that AlphaFold has furnished an answer to that problem. But some of them, including Wolynes, don't think that's an issue, because they believe that problem was satisfactorily solved long before AlphaFold came along. In the eyes of those experts, folding funnel theory is the answer to the why question of protein folding – and it renders the how question moot.

That's because, according to folding funnel theory, the specific pathway by which an amino acid chain folds up into a stable protein doesn't really matter. That might not seem like a particularly satisfying answer, but it's not unreasonable. If protein energy landscapes truly are funnel-shaped then determining the specific mechanism by which any individual amino acid chain folds into its final state might be just as pointless as tracing the trajectory of a droplet of water as it rolls down the bowl of a sink. The possible pathways are innumerable; it's far simpler to focus on the destination.

The best evidence in favor of this perspective, Baker says, is the success that he and others have achieved in the field of protein design. Even in the days before AlphaFold, when physical principles played a key role in the protein design process, Baker didn't concern himself with the specific

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

mechanisms by which a protein he designed would fold. The goal, he said, was simply to design an amino acid sequence that would achieve such an abyssal free energy in its folded state that it couldn't possibly take on any other shape. "We never worry about the pathway," he says.

Baker's reasons may be more practical, and Wolynes's more intellectual, but both are happy to leave the protein folding problem in the past. Rose, though, thinks that there are still important insights left to uncover. Folding funnel theory is appealing, even elegant – but it's also a fairly shallow account of why proteins fold, and it leaves numerous questions unanswered. And for all the folding funnel theory's successes, it still can't accurately predict the structures of many proteins – which, according to what Wolynes argued in his 1987 paper, suggests it might be incomplete.

Rose wants a theory of protein folding that can do both prediction and explanation. The advent of AlphaFold has only motivated him further, because AlphaFold's success demonstrates that protein structures can, in principle, be predicted. The trick is to find a method that humans can actually grasp – and that, unlike AlphaFold, furnishes understanding of how and why proteins fold. That understanding might build on the understanding afforded by folding funnel theory, just as Newton's law of universal gravitation built on Kepler's laws, perhaps by explaining why particular amino acid sequences have funnel-shaped energy landscapes. Or it could replace folding funnel theory altogether, like the Copernican model replaced theories of the solar system that placed Earth at its center.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

Rose is now retired after decades as a Johns Hopkins professor, though he continues to work independently on the protein folding problem. He jokes that he only has one junior researcher working under him now – himself – and that this researcher doesn't always do what he'd like him to. From Rose's perspective, other scientists have turned away from the basic protein folding problem not because it has been solved, or because it's not important, but because it has been around in some form since Pauling and Mirsky published their 1936 paper. Since then, countless scientists have worked to discover a clear, human-usable recipe that explains why particular amino acid sequences adopt certain shapes, but the wait has been discouraging. "People have either given up, or they've concluded that, the truth of the matter is, it doesn't really work that way," Rose says. "I'm convinced that it does."

To achieve the kind of understanding that Rose seeks, scientists have to figure out how to bridge the gulf between the knotty complexity of protein folding – the near infinitude of possible interactions among the hundreds of amino acids that compose the typical protein – and the limitations of human cognition. If we were more intellectually sophisticated, we might be able to understand protein folding using quantum mechanics, which provides a complete account of how the atoms in a protein behave. But that's not an answer that any human can apply to the problem of protein folding. The math is far too complicated, and our brains can only handle so much. Even computers can't solve protein folding with quantum physics.

The goal, then, is to find a theory that maximizes both human usability and predictive utility. But finding that

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

balance is difficult. Folding funnel theory makes sense, but it lacks specificity and can't be used to make predictions; AlphaFold makes great predictions, but it can't yet be understood by humans, at least not with the tools we currently have. Rose thinks it's possible to find a better option.

Not everyone agrees. "When people saw the first proteins, they were gnarlier objects than people expected," Jumper says. "Over many years, there was a lot of understanding that we probably weren't going to get a really simple method of looking at the sequence and a human writing down what the folded structure is." It's entirely possible – maybe even plausible – that protein folding is simply too tricky a process for humans to fully understand. Having a predictive tool as effective as AlphaFold makes that possibility much easier to accept.

Charting the Field's Future

When Eugene Shakhnovich, a biophysicist at Harvard, first started working on protein folding in the 1980s, the contours of the field were defined by Anfinsen's thermodynamic hypothesis and Levinthal's paradox. The central questions concerned the uncanny accuracy and speed with which amino acid chains fold into their proper three-dimensional shapes in test-tube experiments, and the proper tool for answering them was the physics of energy landscapes. Shakhnovich's papers from this era are dense with calculations and formulae; at a quick glance, they look like they might have been published in a mathematics journal.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

These days, however, Shakhnovich's research looks a bit different. He makes regular use of the genetic manipulation techniques afforded by modern molecular biology, and many of his studies include extensive experimental results. Though he still spends some of his time grappling with theoretical puzzles, Shakhnovich has swung the spotlight of his attention toward the biological details of how protein folding happens in the cell.

That's partly because Shakhnovich, like Baker and Wolynes, thinks that some aspects of protein folding are no longer mysterious. "Any field which is successful falls victim to its own success, because many problems get solved and then general interest moves elsewhere." Shakhnovich says. "That definitely was happening to protein folding even before AlphaFold."

Energy landscape theory, Shakhnovich says, does a good job of explaining why proteins fold and suggests that specific folding pathways don't matter much – as long as those proteins are isolated in a test tube. Inside living cells, however, the picture is quite different. Amino acid chains don't form instantaneously; instead, they are forged, link by link, in a large molecular complex known as a ribosome. As the chain emerges, it begins to fold with the aid of the ribosome in a process called cotranslational folding. In some cases, specialized proteins called chaperones assist in the folding process as well. These days, understanding chaperones and cotranslational folding is a focus of Shakhnovich's work.

Studying chaperones allows Shakhnovich to continue to work on the fundamental physical questions about

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

protein folding that first attracted him to the field. In order to do their job right, chaperones must "understand" how protein folding works, and so studying the function of chaperones means studying the process of protein folding. At the same time, chaperones have also expanded the scope of Shakhnovich's work: He is fascinated by the perplexing fact that evolution would create chaperone proteins rather than devising proteins that fold more readily. "I think that this is the protein folding problem of the 2020s," he says.

Shakhnovich is far from the only scientist to have moved from theoretical questions about amino acid chains in test tubes to experimental problems about proteins in cells. Bowman – the biophysicist who defended his field as "Not Dead Yet!" in the wake of CASP 14 – did so after he watched his discipline transform in ways that have nothing to do with AlphaFold. "One of the big changes has been de-emphasis on basic research questions, like how do proteins fold, and way heavier emphasis on problems with more immediate potential translation into therapies or useful technologies," he says. "I've gone and looked for problems that get at the same principles as some of the basic work I used to do, but in a more translational context."

Today, Bowman directs his efforts toward discovering and understanding "cryptic pockets," or hidden spots on the surfaces of proteins that might be promising targets for drugs. Proteins, it turns out, are not rigid, static objects; as they collide with the individual molecules of the fluid that they're suspended in, they twitch and shift, sometimes adopting entirely new shapes. Through investigating these structural variants, Bowman discovers crevices that

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

might be invisible in AlphaFold-predicted structures – and into which small drugs could be able to slip. Studying cryptic pockets allows Bowman to investigate the physical phenomena that undergird protein folding while also working toward disease treatments.

Plenty of protein folding research these days is directed toward medical applications. One major theme is protein *mis*folding, because proteins that fold incorrectly often aggregate into toxic morasses of tangled amino acid chains, which have been implicated in diseases ranging from Alzheimer's to diabetes. Shakhnovich's research addresses misfolding, as chaperones often help prevent it; indeed, preventing misfolding is a common justification for studying protein folding in the cell.

There are numerous reasons why protein folding research has shifted toward problems with clear medical relevance. Some of those reasons are scientific: Many scientists believe that more basic problems in protein folding have been satisfactorily solved, and new technologies have made studying proteins in their cellular context more feasible than it used to be. Some are ethical: By studying cryptic pockets or misfolding, scientists might improve and even save lives from the comfort of their labs. And some of those reasons are financial: Scientists need to keep their labs afloat by obtaining grants, and grant-giving organizations, including public institutions such as the National Institutes of Health, like to fund research with clear practical impacts. According to Bowman, funders' emphasis on practical impacts has intensified over the past couple of decades, and those pressures have reshaped protein folding research.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

There's very good reason for funders to prioritize the practical aspects of scientific research. Protein misfolding-related diseases are devastating, and if protein folding research plays even a small part in developing an effective treatment for Alzheimer's, the human benefit will be immense. But it is worth considering how the priorities of funders might interact with AI to shift the landscape of science. That's an especially pressing concern now, when the Trump administration is working so doggedly to curtail the public dollars available for scientific research in the United States. CASP itself has fallen victim to that campaign: After the competition's NIH grant was not renewed in 2025, Google DeepMind had to step in to keep it afloat for the following year.

As long as advancing our intellectual understanding of protein folding can serve practical goals, scientists interested in such questions need not worry unduly. Shakhnovich and Bowman have successfully built research agendas that allow them to investigate protein folding mechanisms while also attacking problems of clear medical significance. But if advancements in protein folding AI further dissociate intellectual understanding and practical benefits, that window of opportunity could start to close.

When Bowman and I spoke, I asked him how he'd feel if he had access to an AI that could perfectly predict the specific pathway by which any protein folds, as well as its various folded shapes and how it interconverts among them. Using such information, one could find all the druggable cryptic pockets in all the proteins in existence. Would having access to that information satisfy his curiosity, or would he

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD

still want to understand why proteins fold in those ways, and why they adopt those shapes?

"I personally would still want to feel like I had some intuition," Bowman says. "But I would totally understand if there were less research dollars going in that direction."

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:35:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://www.cambridge.org/core/product/7541845E6E7D3B9D29FFBBE3F0C63FCD