Science Will Win | Season 4
Episode 4 | The Future
Final Script

JEREMIAH:
Let's shrink down again, like we did in episode 1, and re-enter the human cell. There's a lot going on in this bustling metropolis. Previously, we paid a visit to the nucleus: you know, the library that holds our genetic code, also known as the instructions for building *you*. Today, we're visiting the ribosome.

The ribosome is like a factory that converts the genetic code it receives from the nucleus into a string of amino acids, like a string of beads. These building blocks don't remain as strings, though. They curl, twist, and fold into three-dimensional structures called proteins — the material that builds, repairs, and consists of the body's most basic and crucial components. And *how* a protein folds determines what its exact job will be in the body.

Different shapes mean different jobs. A protein could be…an enzyme in saliva, breaking down starches and sugars as they enter the body. A protein could also be a hemoglobin carrying oxygen in the blood. Or it could be an antibody, helping to combat diseases in your immune system.

Misfolded or unfolded proteins, as well as those that have been affected by mutations, can potentially lead to jobs not being done, which can manifest as certain diseases in the human body.

Not long ago, the way proteins fold was all but a mystery to scientists. But being able to predict how a particular sequence of amino acids will form in 3D space has the potential to change how new drugs and treatments are developed.

Thanks to breakthrough technologies, scientists know more about proteins, and the way they exist in three dimensions, better than ever before. It's one crucial piece of the drug development puzzle.

Welcome to Science Will Win. I'm your host, Jeremiah Owyang. I'm an entrepreneur, investor, and tech industry analyst. I'm passionate about emerging technologies and the ways they can shape our world.

So far in this season, we've explored how innovations throughout history have brought us to where we are now. We talked about how drug discovery changed from a serendipity-based to data-based endeavor. Then, explored the powerful hardware and

smart software required to accommodate big data. Now, the door to the future of AI in drug discovery is wide open.

Today, for our final episode, we're looking toward the future, to understand where today's advancements could potentially take us.

Only around 12% of drugs successfully progress from phase 1 clinical trials to FDA approval. This is due to traditional methods of drug development that rely on a certain level of serendipity and trial-and-error. But, with newer, more targeted drug development strategies, this has the potential to change.

> *Charlotte:*
> *As the technologies advance that enable us to actually understand the structure of the protein, like understanding your lock, so you could design your key, you can design your therapeutic, uh, molecule, that's been very transformational.*

JEREMIAH:
That's Charlotte Allerton, whom you heard from last episode. She leads preclinical and translational sciences at Pfizer.

Charlotte's work starts with understanding the unmet medical need in patients. And new datasets, and ways of interpreting those data sets, allow Charlotte and her team to predict how a particular treatment might assist in addressing those unmet medical needs.

> *Charlotte:*
> *An explosion in different technologies I think have really opened up our understanding of human biology. Everything from the human genome project to some of the technologies that enable us to assess a number of different measures in cells the different omics measures that we can obtain. And all of that data collectively I think is leading to what is a true transformation in our understanding of human biology.*
>
> *How does that help? Well, I think it opens up different hypotheses on how to treat the disease and from that comes different ideas on the biology we need to moderate in order to be effective at treating the disease. So it translates to me from an explosion in data to a much larger numbers of ideas to treat disease. And I think increasingly a higher confidence in the success of those ideas in truly doing what we are hoping they're gonna do when we get into clinical development and ensuring that we truly are treating the disease.*

JEREMIAH:

In the early days, if Charlotte wanted to know whether or not a particular molecule or compound would make a potential new drug, she'd have to make it in a lab, testing to see how it interacted with proteins in the human body, and measuring all of its properties manually. That is a lot of time consuming labor.

> *Charlotte:*
> *And now what I see is, uh, scientists are looking at all of the molecules they could make on a computer and then applying some of these machine learning and AI algorithms to inform them, okay, of all these molecules I could make, which are the ones that are most likely to be a clinical candidate, and then making those and, um, testing those in the laboratory.*

JEREMIAH:

AI can assist in a process called virtual screening, that allows Charlotte and her team to understand a potential molecule before even making it. It helps them prioritize which molecules are most likely to become good medicines.

> *Charlotte:*
> *Early in the process try and understand do we have a, a molecule that can block a certain protein? And sometimes we take the protein and we screen lots of molecules and we test lots of molecules and we see which block it at. Other times we can do that whole process virtually and we can have a predicted structure of the protein or potentially an actual structure of the protein, and then we can look at literally billions of potential molecules that we could make to see which computationally would be predicted to bind to that particular protein. And you can imagine doing that computationally is very efficient compared with actually running those experiments over millions of compounds in a laboratory and then looking at the data and deciding which ones, uh, should you follow up on.*

JEREMIAH:

Predicting what molecules become good medicines requires predicting how those molecules will bind to proteins in the human body. And *that* can be helped along by being able to predict what the protein looks like. Essentially, making medicine is all about binding the right molecule to the right spot on the right protein, in order to modulate how that protein functions. So suffice it to say, understanding proteins is a big deal. Here's Daniel Ziemek again from last episode, Vice President of Integrative Biology and Systems Immunology at Pfizer.

> *Daniel:*
> *And at some point, the time comes where you want to either develop a, a chemical molecule or a so-called antibody, which is a little bit of a bigger*

molecule that has certain properties that you wanna then use to, to inhibit the function of the protein you have selected. Right? So which of the many molecular machines that are doing their thing in your body right now do you wanna change to cure the disease you have? Right? Because presumably one of these little molecular machines, proteins, are not working the way they're supposed to. And one of the big things we usually do is we take out a specific one of these molecular machines that is not doing the right thing and that usually helps with the disease you have and kind of obviously, um, it, this works better if you know what this machine looks like

JEREMIAH:
Since the 1960s, scientists have been trying to figure out how to predict what shape proteins will form based on their amino acid sequences. The folding process is *very complex,* so for decades, this question remained unanswered.

Daniel:
It's interesting that, that we still don't know all the structures, um, of all the proteins in, in humans or, or in other animals.

JEREMIAH:
What scientists *do* know is that proteins are made up of long chains of amino acids, which are chugged out by the ribosome. And each amino acid is determined by a unique genetic code — a set of rules turning nucleotides in DNA and RNA into proteins. Genetic code is made up of endless combinations of the letters A, C, G, and T – which stand for the different nucleotide bases: adenine, cytosine, guanine, and thymine.

Daniel:
And one of the breakthrough happened when humans figured out how to read, whether the next position in this long, long word is an A, C, G, or T, and all of a sudden we could take bits and pieces of this long word and say, aha, it's A-AC-C-G-G-T-T-T-T-C-C-C, right? And that was a big breakthrough in, in so-called sequencing technology that has revolutionized a lot about what data we have at our disposal to make progress.

JEREMIAH:
Decoding the order of the nucleotides in genetic code was one important breakthrough in biology. Depending on the configuration of code that comprises a protein, it will fold into a unique 3D structure.

Daniel:

Proteins have to arrange themselves in three dimensional space and lots of physical forces come together to determine the ultimate shape. And that is not a simple algorithm.

JEREMIAH:
In 2020, huge progress was made on this front. A team of researchers at DeepMind successfully created a model called AlphaFold that could predict a protein's final end shape. This model helped answer one of the holy grail questions of biology: how does a long line of amino acids configure itself into a 3D structure that becomes the building block of life itself?

In October 2024, two scientists from DeepMind and a third from the University of Washington won a Nobel Prize for these huge advancements in protein structure prediction.

Daniel:
these AI mechanisms, these deep learning technologies seem to be able to find shortcuts that lets us quickly get from the beginning, from this sequence, from this long string to the final shape of the protein. So we do not have to simulate every little piece, which we could do in theory, but can't do so well in practice, and we find a shortcut that lets us jump to the conclusion.

JEREMIAH:
Daniel says the emergence of AlphaFold is due in part to the technology advancements that came before it: the availability of big data, as well as the machine learning models like AlexNet that could identify cat photos, and even ChatGPT.

Daniel:
All these different fields have used similar techniques and relied on availability of, of computational power, and that has sort of paved the way that a lot of the building blocks were there and they could be translated and tried out in this protein folding space.

JEREMIAH:
AlphaFold was made available for free, and DeepMind built an open database that provides access to 200 million protein structure predictions. It also regularly updates with structures for newly discovered protein sequences. So far, the AlphaFold database has over two million users in 190 countries — saving millions of dollars and millions of years in research time.

Daniel:

AlphaFold woke people up to the possibilities that are there with this new AI technologies in this field. And this is really starting to come into pharmaceutical companies like Pfizer to directly, impact how we think about protein structure prediction, and then what we know about when we develop these, these, compound, we say these chemical molecules that can become future medicines or antibodies and where they bind to these proteins, because that enables us to do that much better than we used to do,

JEREMIAH:
Beyond understanding the 3D shapes within cells, it's also important to drug design to understand where cells exist within tissues.

Now, Dr. Raza Ali is a group leader at the University of Cambridge Cancer Research UK Institute. He's also a pathologist. Basically, that means he looks at human tissue under a microscope to determine whether it's normal, abnormal, or cancerous. But earlier in his career, Dr. Raza Ali was working at a lab in Zurich, Switzerland that invented a process that would become key to pharmaceutical science: spatial imaging.

Raza:
Up until that point, we'd really thought of cancers in two ways. Either it's using very artificial and constrained models in the lab, namely, or immortalized cancer cell lines, which can, you can grow in a dish. And so that's a very limited view on the totality of human cancer

JEREMIAH:
And this method was successful for a long time.

Raza:
but pathologists had, you know, as I say, for over a hundred years, thought of tumors as, um, caricatures of normal tissue. So we look at in space how these, how these tumors have put, you know, disturbed normality and, and become very bizarre, um, representations of, of normal tissue. And, and so that involves the deviation from normality involves the interaction with, um, with normal tissues and normal, the normal immune response and so on.

JEREMIAH:
But the lab Raza worked at realized there was a new frontier in cancer research. That frontier is spatial omics.

Raza:
Well, in fact, it's something of a misnomer or a rebranding, if you like, because essentially all biology is spatial, all occurs in a spatial context. Okay. Nothing

occurs totally isolated and dissociated from its surrounding tissue or surrounding environments. but the reason it's come to the fore now is because a suite of technologies in the past 10 years have discovered and commercialized, which allow us to characterize particularly tissues, but also other settings under the microscope in much more detail than we have in the past.

JEREMIAH:
You might remember in episode 1, when Kellie Kravarik, Cellular Genomics Group Leader for Systems Immunology at Pfizer, talked about measuring the smoothies of biology.

> *Kellie:*
> *You would take samples from a tissue culture experiment or a human patient and, and you would blend them up and you would measure all of the blended components.*

JEREMIAH:
If bulk RNA sequencing, like Kellie was talking about, is a fruit smoothie, spatial omics is like a fruit tart — with each fruit individually placed in a particular configuration. A circle of blueberries around the edge, kiwi and mango fanned out in slices, a crown of strawberries in the center. The arrangement of the fruit is just as important as each individual piece of fruit itself. The same goes for cells — the arrangement of particular cells can be very informative when characterizing a tumor's microenvironment. Here's Raza:

> *Raza:*
> *Nowadays we can see very high fidelity segmentation of single cells, for example, which previously hadn't been possible. That's enabling us to probe the biology of tissues, and particularly in biology of cancer with much greater, greater precision. The end result is a highly, um, quantitative detailed map of each image, which tells us about all the different cells contained in that, in that image, what their expression profiles are with respect to the antibodies we selected, and also what their morphology are.*

JEREMIAH:
Here's how spatial imaging works.

The immune system partially consists of antibodies, whose job it is to bind to antigens and destroy them. The spot on the antigen where the antibody can bind is called the epitope. Raza and his team conduct epitope-based imaging, which allows them to characterize around 44-55 proteins within a particular section of tissue.

Raza:

So a very thin piece of tissue, which essentially looks transparent and is sitting on a glass slide, is treated with all of these antibodies simultaneously. So those antibodies will then go and chase down the protein of interest that they target and stick to it. Um, and so we wash off all of the excess. And that leaves us with a, a piece of tissue, which contains lots of antibodies stuck to very specific, epitopes

JEREMIAH:

The antibodies, attached to the antigens via the epitopes, are tagged with metals. The antibodies are then ionized through a technique called imaging mass cytometry. This allows them to be counted more precisely.

Raza:

if you think of an old style television picture is made, made up of a, uh, cathode ray, sort of rastering across the screen.

JEREMIAH:

Ever wondered how old TVs work, or why a TV was sometimes called "a tube?" That's due to the cathode ray tube inside, which was heated so that it emitted electrons. Then, the electrons beamed something called a raster pattern onto a glass surface to create the pixels that make up moving images.

Raza:

We essentially do something very similar except where the raster goes across the screen each time it's returning values across these 44 proteins. And so we have a stack of 44 images simultaneously for a given tissue. And that's essentially the data we then go on to process and try to understand.We subject those images to image analysis, which is highly automated, where we, uh, essentially segment the image very precisely so we know what different regions contain, particularly at the single cell level.

JEREMIAH:

Once they have these highly complex images, Raza and his team will look at different aspects of them, depending on the experiment.

Raza:

But often they are things which are associated with disease outcome. So whether somebody is at high risk of death or a low risk of death, or relapse, or with treatment response. So whether a given tumor is going to respond to a, a given treatment or not. And this technique has been shown to be particularly useful for looking at responses to treatments like immunotherapy, which are a relatively recent addition to the armamentarium, in breast cancer, for example.

JEREMIAH:

From deep neural networks predicting protein folding structure to virtual spatial omics screening assisting in cancer research, implementing AI in drug research, discovery, and development involves several strategic steps.

According to Enoch Huang, whom you've heard from throughout this season, there are four key areas where the field of pharmaceutical research and discovery have been particularly successful in developing and applying artificial intelligence. Those four areas are training data, infrastructure, computation algorithms, and practitioners.

Enoch likens these four things to a car driving on a road. First is data, but not just any data, data that is processed with machine learning in mind.

> *Enoch:*
> *So the analogy here is around fuel, right? Because you don't pump crude oil that you extract from the ground into your car. It needs to be processed in refineries to be gasoline or petrol, right? So this both the, the raw data as well as people who know how to process the data as input into machine learning models.*

JEREMIAH:

Next is infrastructure.

> *Enoch:*
> *So it's databases, it's software, it's computational resources. Here are the roads, the bridges and tunnels necessary for machine learning to happen at scale because you, you know, the, the cars, the drivers, they need to go places. They need. We need to be able to fuel our cars in pumping stations. We need to be able to, um, make the models usable.*

JEREMIAH:

Then there's access, adaptation and development of modern ML algorithms.

> *Enoch:*
> *So here is, you know, the brain power and experts who understand the possibilities of machine learning developments in the broader field, but thinking how to adapt them to the problems that are germane for pharmaceutical R&D. And so here the product is sort of a more powerful engine that resides in the car.*

JEREMIAH:

Then comes the fourth, and maybe most important, component: the humans working collaboratively to drive all the other components forward. He refers to them as the skilled drivers.

> *Enoch:*
> *Here you have computational practitioners that can use all this infrastructure, the tools, the powerful engines to drive projects. And so you need to have practitioners who understand the biology and the chemistry and the potential for these algorithms, these methods for the benefit of their portfolio.*

JEREMIAH:
Marrying data, hardware, and software is what enables the newest innovations in drug discovery and personalized medicine through AI. Despite challenges, the use of AI in this field is primed for continued growth. As AI becomes more embedded into everyday research, we may only be at the tip of the iceberg. Here's Daniel again.

> *Daniel:*
> *Before you would have had to have a lot of money to find out the specific protein structure and then go from there. Now you can predict all these different structures and, and look through them and, and pick the one that that works best. So it I think opens the mind to different approaches of how we do drug discovery in specific areas. And that is still growing. I don't think we have understood how to best use a lot of these new AI methods, fully.*

JEREMIAH:
The potential of these new technologies is only just emerging.

> *Charlotte:*
> *If I was to fast forward to the future and say, what would I really like to be able to do, I think certainly the interrogation of human biology and predictive models based on the wealth of omics data that we already and will have more of available by then will be very enabling in everything from selecting the right proteins to modulate in the body, through to selecting the right patients to put into our clinical studies to determine whether a particular therapeutic, uh, works well or not. I think we're only just scratching the surface*

> *Raza:*
> *We're still only at the early phases of really understanding what's possible.*

> *Daniel:*
> *The pace of innovation is not slowing down as far as I can tell right now. So I think we'll have much, much deeper insight into, uh, the pathogenesis of disease.*

So why do disease happen? And we will get more insights into how can we fix it with all these technologies. It's amazing. I think we live in the future.

JEREMIAH:

That's all for this season of Science Will Win. Thank you so much for joining me on this journey. We hope you join us for future seasons!

Science Will Win is created by Pfizer and hosted by me, Jeremiah Owyang. It's produced by Wonder Media Network. Please take a minute to rate, review and follow Science Will Win wherever you get your podcasts. It helps new listeners to find the show.

Special thanks to all of our guests and the Pfizer research and development teams. And thank you for listening!