Harnessing Artificial Intelligence to Fight Infectious Diseases

In 1928, Sir Alexander Fleming uncovered penicillin, an antibacterial compound that would alter the course of medicine. By the 1940s, this miracle drug entered clinical use, and humanity began to rely on penicillin and other antibiotics to treat once-lethal bacterial infections. Despite this success, bacteria continue to threaten global health as antibiotic-resistant strains emerge and spread. Scientists now race against time to develop new antimicrobial drugs before bacteria gain the upper hand. Fortunately, researchers have found a powerful ally: artificial intelligence.

Welcome to *The Scientist* Speaks, a podcast produced by *The Scientist'*s Creative Services Team. Our podcast is by scientists and for scientists. Once a month, we bring you the stories behind news-worthy molecular biology research.

In this episode, Charlene Lancaster from *The Scientist* spoke with César de la Fuente, a presidential associate professor at the University of Pennsylvania, to learn how his team is leveraging AI to discover novel antibiotics from unique sources, including spiders, archaea, woolly mammoths, and ancient and modern humans.

<u>The Looming Threat of Drug-Resistant Bacterial Pathogens</u>

Narrator:

As a child, César de la Fuente was always fascinated by the secrets of life and nature. He spent hours with his siblings exploring, experimenting, and asking questions about how living organisms function. That curiosity guided him toward a PhD focused on bacteria, the simplest lifeforms, and on the smallest functional biomolecules, short polypeptide chains. Bacteria captivated him because they are among the original organisms on the planet. They have existed since the beginning of life, constantly adapting and surviving across immense spans of evolutionary time. Studying them offered insights not only into their biology but also into the strategies organisms use to endure. These conceptual threads continue to shape the work he does with his team today, developing new antibiotics to combat the rise of drug-resistant bacterial pathogens.

César de la Fuente:

Bacteria are becoming increasingly resistant to every antibiotic drug that we have available. Currently, these infections are associated with about five million deaths per year in the world. By 2050, that number is predicted to double to ten million deaths per year around the world. If you run a quick calculation, that is about one death every three seconds. So, I think this is the greatest existential threat to humanity. And if you study human history, infectious diseases are the number one killer of humans. We now happen to have antibiotics, but imagine a post-antibiotic world, where these miracle drugs no longer work. We could be back to the pre-penicillin era, where the life expectancy was about half of what it is now. And you would be playing in the playground, fall, and that simple wound could actually get infected and that could be a lethal situation. So, I think modern medicine really collapses without antibiotics, and communicating the value of these medicines to the public I think is essential. But of course, also coming up with new solutions before it is too late, and arguably it is already too late.

Narrator:

For less than a century, antibiotics have stood at the center of modern medicine, enabling routine medical interventions such as childbirth, organ transplantation, and chemotherapy treatment. Yet during that short time, their discovery has slowed to a crawl. Over the past decade, only a handful of structurally novel antibiotics have reached the market.

Traditionally, discovering new antibiotics was a physical and painstaking process. Scientists traveled the globe, digging into soil, sampling sediments, and collecting water from diverse environments in search of promising molecules. This method relied heavily on trial and error and often yielded few useful candidates.

The search for new antibiotics remained slow, uncertain, and labor-intensive, leaving researchers to wonder whether there might be a better way.

The Potential of AI for Antibiotics Discovery

César de la Fuente:

Towards the end of my PhD, I sort of had this epiphany. It became really clear to me that with increasing amounts of data in biology and increasing computing power, we were in an ideal position to use AI and to use machine learning to try to apply it to something as complex as biology, and more specifically, to microbiology and to antibiotic discovery. So then, I got recruited at MIT after my PhD. MIT at the time was probably the Mecca in the world for AI research. But when I got there, most people were applying machine learning systems to recognize voices or faces. So, pattern recognition algorithms. But people were not really applying AI to biology. When I proposed that I wanted to train a machine to create a new antibiotic, the reception that I received was one of skepticism. Most people thought that this was a crazy idea because biology has too many variables. And most people would tell me that an algorithm would not be helpful because you are dealing with a level of complexity that cannot be captured by AI. Maybe because I was younger at the time, along with my collaborators, we ignored that skepticism. We really in our guts believed that it would be possible to train a machine to create from scratch a new antibiotic. We persisted, and we were able to evolve synthetic molecules that when we synthesize them in the lab, we were actually capable of killing bacterial pathogens that kill many, many people every single year in the world. Ultimately, we tested some of those compounds in mouse models of infection, and they were able to reduce the infection. That was sort of the aha moment. We can design new molecules on the computer that when we synthesized them, they are actually capable of reducing infections in mice, right? And that really convinced us that this could be a whole new field of research.

Narrator:

This work produced the first computer-discovered antibiotic effective in animal models. Central to this achievement was AI, which refers to machine systems capable of mimicking human intelligence. Machine learning algorithms fall under this umbrella, and people design these models to learn specific tasks that interest humans but would be extremely difficult for the human mind to tackle alone. Typically, these tasks involve enormous amounts of data, such as examining thousands of genomes to uncover hidden patterns. What might take a scientist decades to analyze, a machine can now complete in just a few hours. In this way, AI acts like a powerful accelerator, much like the calculator once did, only far more capable, allowing de la Fuente and his team to explore biology in completely new ways.

César de la Fuente:

One of the conceptual ideas that we had a number of years ago was to think of all biology as an information source. Because if you think of biology as just a bunch of code, then you can start to devise algorithms that can explore all that code to try to find hidden patterns that correspond to potential antibiotic molecules. They are really hidden from the human visual cortex and from our ability to actually identify them. But with a lot of the computational methods that we have developed over the years, we can now systematically and digitally identify antibiotics in a few minutes or hours instead of having to wait for years.

Narrator:

This idea inspired de la Fuente and his group to look for antibacterial compounds close to home, within the human proteome. This was the first time scientists ever explored this space as a source of antibiotics. The team found thousands of new antimicrobial molecules encoded in the human proteome. They called these amino acid patterns encrypted peptides because of their small size and hidden location within larger proteins, most of which were not previously known to have antibacterial activity or roles in the immune system. This discovery led the team to consider other potential sources of antibiotics.

Using AI to Discover Antimicrobials in Extinct Organisms

César de la Fuente:

That work really sparked a lot of new ideas in my lab, one of which was most likely these compounds are not just confined to the human proteome. And we had this idea that maybe they would be found throughout evolution and across the tree of life. So, we did a whole research project on Neanderthals and Denisovans, and we found antimicrobial molecules in those ancient humans as well. And that was the moment when it really convinced us that it was possible to do this molecular de-extinction work. So, we use machine learning to discover functional biomolecules. So, we focus on small peptides and small proteins, which are the workhorses of life. They are the nanomachines that do everything in nature. We see this as a conceptual framework to identify biofunctional molecules throughout evolution and then synthesize them in the lab to be able to learn how changes or mutations that occurred over time in those biomolecules influence their biological function. And that function can be antimicrobial properties, but it can also be an ability to modulate the immune response. It can be anti-cancer properties.

Narrator:

The work with Neanderthals and Denisovans inspired the team to explore biology on a much larger scale. They decided to sample all the ancient biological data available in the world. However, they quickly realized that the project would create a huge scaling challenge. The group normally analyzed the proteomic data of a few organisms, but studying all of ancient life required them to assess data from hundreds of organisms. To overcome this challenge, they developed a powerful deep learning model called antibiotic peptide de-extinction, or APEX. This sequence-to-function prediction model examines amino acid sequences and predicts which peptides are likely to have antibiotic function. Using APEX, the team rapidly identified promising molecules from a vast range of ancient organisms.

César de la Fuente:

APEX was wonderful because it really opened a window into the past. It enabled us to sample the genetic information of organisms all throughout evolution, and it took us through this journey through the Holocene and the Pleistocene. It enabled us to identify new antibiotic molecules in organisms that were unimaginable, like ancient penguins that were extinct in the 50s, magnolia trees that disappeared throughout evolution, ancient elephants, woolly mammoths, giant sloths, and many other organisms that used to roam and used to exist on our beautiful planet. We think of this as a way of using novel technologies like AI to learn from a lot of these molecular fossils or molecular relics that were honestly lost, right? We believe there is a lot of intelligence that can be extracted in biology by studying a lot of this ancient biological data.

Hunting Antibiotics in Nature's Most Unusual Sources

Narrator:

His team also used APEX to discover novel antibiotics within animal venoms. Venoms are the product of millions of years of evolution, but scientists still do not fully know their chemical composition. The team focused on identifying small peptides in venoms that had predicted antibiotic function. APEX allowed them to digitally screen all the venoms available worldwide, revealing molecules with antibacterial properties. These findings suggest that the venom-producing organisms, which include snakes, spiders, and cone snails, use these complex mixtures both to kill competitors and potentially to protect themselves from infections. In this way, venoms act like a Swiss army knife for host defense.

Most recently, de la Fuente and his group retrained APEX with additional experimental data, creating an improved version called APEX 1.1. They used this upgraded model to explore another branch of the tree of life for antibiotics: archaea.

César de la Fuente:

Archaea are these very esoteric organisms that we really knew very little about. For the longest time, scientists did not really know what to make of archaea. We are still puzzled by their amazing ability to survive in environments that would kill a human within minutes. Archaea can live in volcanoes, in environmental conditions of extreme pH and temperature that would be extremely lethal to humans. And the other thing that I find fascinating about archaea is that they are one of the initial colonizers of Earth, one of the first living forms to have ever existed on our planet. So, we can learn a lot from them.

We found what we call archaeasins that are encoded in archaeal proteomes. These are peptide molecules that can kill some of the nastiest pathogens in our society, and a couple of them were effective in mouse models, including archaeasin-73, which was comparable to last-resort antibiotics like polymyxin B in its ability to reduce infections in mouse models.

From an evolutionary perspective, perhaps that tells us that they had bacteria as competitors in this initial Earth when life first originated on our planet. Perhaps archaea were producing these antibiotics to counter bacterial competitors. Obviously, that is only speculation. But through all these findings and all this work that we are doing, enabled by AI, and by a lot of experiments, you know I think we can unlock new questions, new ideas that we can test in the lab, and these are some of those ideas.

Developing Advanced AI Models

Narrator:

Most existing AI models for antibiotic discovery are strain-specific and perform well only on the bacterial species used to train them, preventing their use for microbes that lack antimicrobial training data. To address this limitation, de la Fuente and his team developed a new AI model called ApexOracle, which they are currently validating in the laboratory. ApexOracle analyzes a bacterial pathogen's genome to predict which existing antibacterial drugs are likely to be effective against the microbe. In addition to its prediction capabilities, the model can also function as an antibiotic designer.

César de la Fuente:

ApexOracle also has a generative AI component, where it can generate novel small molecules and new peptides that are predicted to be active against that genome. For the prediction task, it takes seconds. For the generation tasks, it takes minutes, maybe five minutes. So, it is super quick. I find it exciting because if it works experimentally, it can enable us to keep pace with resistance. As new clones emerge, we might be able to create new molecules that are effective against those clones.

Narrator:

Just as de la Fuente's group is building AI models that can perform more than one task, they are also applying this concept to drug discovery. Their latest tool, APEX DUO, generates molecules with dual functions rather than a single activity. With this approach, the team aims to tackle one of the toughest challenges in medicine: intracellular bacterial infections.

César de la Fuente:

Bacteria penetrate into our human cells, and they hide inside the cells in such a way that the immune system cannot really kill them. Conventional antimicrobial therapy cannot target them either because conventional antibiotics cannot penetrate into our cells. These bacteria create these intracellular infections that are really, really hard to treat. We came up with this generative AI model that can generate peptide molecules that are predicted to have two functions. This is kind of venturing into this concept that we call multimodality, where these peptides are capable of penetrating into human cells, and then once inside the cell, they are capable of killing the bacteria. This is a very hard problem because you have to design molecules that are capable of doing two tasks in biology. Once we synthesized the compounds that the computer generated, we saw that the vast majority of them were capable of killing bacteria, but only a few of them were capable of penetrating into human cells. We were anticipating this because this was the first time that anybody had attempted this. This is a proof of principle of the concept of designing multimodal molecules.

Building Datasets Through Teamwork and Perseverance

Narrator:

Designing dual-function molecules demonstrates the power of AI to tackle complex biological problems. These advances show what is possible, but applying AI successfully in microbiology requires high-quality experimental datasets to train the predictive and generative models. Producing such datasets is one of the major challenges of using AI as a tool in biology, because without reliable and consistent data, even the most sophisticated algorithms cannot produce meaningful predictions or identify promising new molecules. Generating these datasets requires careful experimentation and analysis, often taking months or years of laborious work.

César de la Fuente:

When I was recruited at UPenn, I bet really strongly on generating our own datasets. From the beginning, we designed them. We always use the same conditions, the same pH, the same temperature, the same bacterial strains. We try to decrease the degrees of freedom in biology, which is extremely difficult. When we started building this dataset, I had this calculation that I thought it would take us about two and a half years to generate enough data experimentally to then train a state-of-the-art AI model. But that turned out not to be the case. After two and a half years, we tried to train the model, and it was not successful. So then, there was a moment of huge uncertainty, right, in the lab. I had to convince my team to keep going because I believed that this was the only way to really be able to tackle biology head-on by having really good datasets that we created from scratch. And it took us an additional year. It took us three and a half years and amazing people in the lab that were willing to walk through walls and walk through that uncertainty to be able to achieve this dream. After three and a half years, we had enough data, we trained the model, and it was successful for the first time.

I am very proud of the team because we were willing to jump into the abyss of knowledge. I am really lucky to be able to work with some of the smartest people in the world, and they come from completely different backgrounds. They think very differently about problems. Right now, we have chemists, physicists, computer scientists, microbiologists, biologists, engineers, working together. I am just really lucky to be able to brainstorm with them, come up with entirely new, crazy ideas to problems. And I think that is the power of bringing people together that think differently and creating an environment where everybody can just speak their mind and try to work together collaboratively.

We believe that just my lab and a couple other labs working on this is not sufficient. We are going to need the whole scientific community to come together to actually tackle such a big problem as antimicrobial resistance or AMR.

Outro

Thank you for listening to *The Scientist* Speaks. This episode was produced by the Creative Services Team for *The Scientist* and narrated by Charlene Lancaster. Please join us again in December, as we learn about the Space Omics and Medical Atlas. To keep up to date with this podcast, follow *The Scientist* on social media and subscribe to *The Scientist* Speaks wherever you get your podcasts.