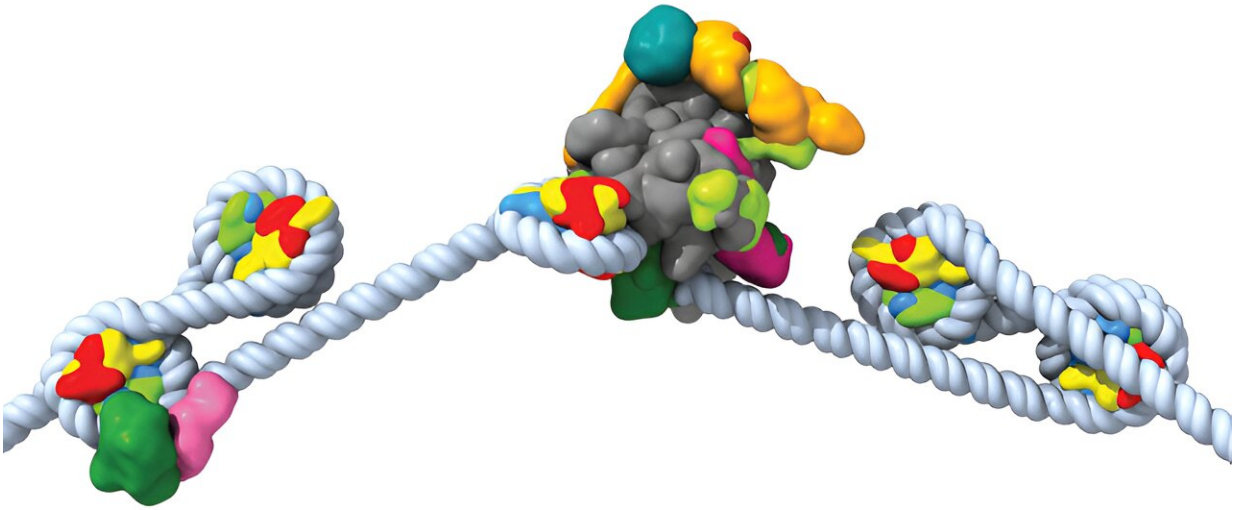
Q&A: How machine learning is propelling structural biology

July 22 2024, by Catherine Caruso



Molecular machines, a chromatin remodeler (pink and green on left) and RNA polymerase II (gray, yellow, and blue in center), work together to read genomic information stored on tightly packed DNA (white coil). Credit: Farnung lab

For Lucas Farnung, there is no question more fascinating than how a single fertilized egg develops into a fully-functioning human. As a structural biologist, he is studying this process on the smallest scale: the

trillions of atoms that must synchronize their work to make it happen.

"I don't see a big difference between solving a 5,000-piece jigsaw puzzle and the research we are doing in my lab," says Farnung, an assistant professor of cell biology in the Blavatnik Institute at Harvard Medical School. "We are trying to figure out what this process looks like visually, and from there we can form ideas about how it works."

Nearly all cells in the human body contain the same genetic material, but what tissue types those cells become during development—whether they become liver or skin, for example—is largely driven by gene expression, which dictates which genes are turned on and off. Gene expression is regulated by a process called transcription—the focus of Farnung's work.

During transcription, molecular machines read instructions contained in the genetic blueprint stored inside DNA, and create RNA, the molecule that carries out the instructions. Other molecular machines read RNA and use this information to make proteins that fuel almost all activities in the body.

Farnung studies the structure and function of the molecular machines responsible for transcription.

In a conversation with Harvard Medicine News, Farnung discussed his work and how machine learning is accelerating research in his field.

What is the central question your research seeks to answer?

I always say, we are interested in the smallest logistical problem there is. The human genome is present in almost every cell, and if you stretched out the DNA that makes up the genome, it would be roughly two meters,

or six and a half feet long. But this two-meter-long molecule has to fit inside the nucleus of a cell, which is only a few microns in size. This is the equivalent of taking a fishing line that stretches from Boston to New Haven, Connecticut, or about 150 miles, and trying to squeeze it into a soccer ball.

To achieve this, our cells compact DNA into a structure called chromatin, but then molecular machines can no longer access the genomic information on DNA. This creates a conflict, because DNA needs to be compact enough to fit inside a cell's nucleus, but molecular machines have to be able to access the genomic information on DNA. We are especially interested in visualizing the process of how a molecular machine called RNA polymerase II gains access to genomic information and transcribes DNA into RNA.

What techniques do you use to visualize molecular machines?

Our general approach is to isolate molecular machines from cells and look at them using specific types of microscopes or X-ray beams. To do this, we introduce genetic material that codes for a human molecular machine of interest into an insect or bacterial cell, so the cell makes a lot of that machine. Then, we use purification techniques to separate the machine from the cell so we can study it in isolation.

However, it gets complicated because often we are not just interested in a single molecular machine, which we also refer to as a protein. There are thousands of proteins that interact with each other to regulate transcription, so we have to repeat this process thousands of times to understand these protein-protein interactions.

Artificial intelligence is starting to permeate many

facets of basic biology. Is it changing the way you do structural biology research?

For the last 30 or 40 years, research in my field has been a tedious process. A Ph.D. student's career would be dedicated to learning a little bit about a single protein, and it would take thousands of students' careers to learn about how proteins interact in a cell. However, over the last two or three years, we are increasingly looking to computational approaches to predict protein interactions.

There was a big breakthrough when Google DeepMind released AlphaFold, a machine-learning model that can predict protein folding. Importantly, how proteins fold determines their function and interactions. We are now using artificial intelligence to predict tens of thousands of protein-protein interactions, many of which have never been experimentally described before. Not all of these interactions are actually happening inside cells, but we can validate them with lab experiments.

This is super exciting because it really accelerates our science. When I look back at my Ph.D., the first three years were essentially a failure—I wasn't able to find any protein-protein interactions. Now, with these computational predictions, a Ph.D. student or postdoc in my lab can be pretty confident that a lab experiment to validate a protein-protein interaction is going to work. I call it molecular biology on steroids—but legal —because we can now reach the actual question we want to answer much quicker.

In addition to efficiency and speed, how else is AI reshaping your field?

One exciting change is that we can now, in a nonbiased way, test any

protein in the human body against any other protein to see if they could potentially interact. Machine-learning tools in our field are causing disruption similar to the disruption to society caused by personal computers.

When I first became a researcher, people were using X-ray crystallography to reveal the structure of individual proteins—a beautiful, high-resolution technique that can take many years. Then, during my Ph.D. and postdoc, cryo-electron microscopy, or cryo-EM emerged—a technique that allows us to look at larger and more dynamic protein complexes in high resolution. Cryo-EM has enabled a lot of progress in our understanding of biology over the past 10 years and has sped up drug development.

I thought I was lucky to be part of the so-called resolution revolution brought about by cryo-EM. But now, it feels like machine learning for protein prediction is bringing a second revolution, which is just amazing to me, and makes me wonder how much more acceleration we are going to see.

In my estimate, we can probably now do research five to 10 times faster than we could 10 years ago. It will be interesting to see how machine learning transforms how we do biological research in the next 10 years. Of course, we have to be careful about how we manage these tools, but I find it exciting that I could make findings on problems I've thought about for a long time 10 times faster.

What are the downstream applications of your work beyond the lab?

We are learning about how biology works in the human body on a basic level, but there's always the promise that understanding basic biological

mechanisms can help us develop effective treatments for various conditions. For example, it turns out that the disruption of the DNA-chromatin structure by molecular machines is one of the main drivers of many cancers. Once we figure out the structure of these molecular machines, we can understand the effect of changing a few atoms to replicate mutations that would lead to cancer, at which point we can start to design drugs to target the proteins.

We just started a project in collaboration with the HMS Therapeutics Initiative that is looking at a chromatin remodeler, a protein that is heavily mutated in prostate cancer. We recently obtained the structure of this protein and are performing virtual screens to see what chemical compounds bind to it. The hope is that we can design a compound that inhibits the protein, and has the potential to be developed into a full-fledged drug that might slow the progression of prostate cancer.

We are also studying proteins involved in neurodevelopmental disorders such as autism. This is a place where machine learning can help us, because the tools we are using to predict protein structures and protein-protein interactions can also predict how small-molecule compounds will bind to proteins.

Speaking of collaboration, how is working across research areas and disciplines important for your research?

Collaboration is super important for my research. The biology landscape has become so complex with so many different research niches that it's impossible to understand everything. Collaboration allows us to get people with different expertise together to work on important biological problems, such as how molecular machines access the human genome.

We collaborate with other researchers at HMS on many different levels. Sometimes, we use our structural expertise to support the work of other labs. Other times, we have solved the structure of a certain protein, but we need to collaborate to understand the role of that protein in the broader cellular context. We also collaborate with labs using other types of molecular biology approaches. Collaboration is really fundamental to drive progress and better understand biology.

Provided by Harvard Medical School

Citation: Q&A: How machine learning is propelling structural biology (2024, July 22) retrieved 22 July 2024 from https://phys.org/news/2024-07-qa-machine-propelling-biology.html