

Predicting the Structures of Proteins

Kathryn Tunyasuvunakool fell in love with computer programming while studying roundworms. Today she uses that passion to help predict protein structures.

By Katherine Wright

Kathryn Tunyasuvunakool grew up surrounded by scientific activities carried out at home by her mother—who went to university a few years after Tunyasuvunakool was born. One day a pendulum hung from a ceiling in her family’s home, Tunyasuvunakool’s mother standing next to it, timing the swings for a science assignment. Another day, fossil samples littered the dining table, her mother scrutinizing their patterns for a report. This early exposure to science imbued Tunyasuvunakool with the idea that science was fun and that having a career in science was an attainable goal. “From early on I was desperate to go to university and be a scientist,” she says.

Tunyasuvunakool fulfilled that ambition, studying math as an undergraduate, and computational biology as a graduate

student. During her PhD work she helped create a model that captured various elements of the development of a soil-inhabiting roundworm called *Caenorhabditis elegans*, a popular organism for both biologists and physicists to study. She also developed a love for programming, which, she says, lent itself naturally to a jump into software engineering. Today Tunyasuvunakool is part of the team behind DeepMind’s AlphaFold—a protein-structure-prediction tool. *Physics Magazine* spoke to her to find out more about this software, which recently won two of its makers a Breakthrough Prize, and about why she’s excited for the potential discoveries it could enable.

All interviews are edited for brevity and clarity.

What is AlphaFold and what can it be used for?

AlphaFold is a machine-learning model that can predict a protein’s structure from its amino-acid sequence. Protein sequences are relatively easy to obtain, with many experiments now able to quickly determine a given protein’s 1D amino-acid chain. But this sequence doesn’t explain how the protein will fold up into a 3D structure, which determines how the protein functions. Folded structures can be experimentally obtained but doing so is time consuming. AlphaFold can predict the structures in a fraction of the time, accelerating the understanding of these systems.

What is your role on the AlphaFold team?

When I first joined the team, I worked as a software engineer, writing data pipelines that take existing experimental protein-structure data and turn them into features we can use



Credit: K. Tunyasuvunakool

to train the model. While doing that, I became really interested in how useful AlphaFold's predictions were. I started to scrutinize the predictions, performing detailed comparisons with literature findings. I then moved into doing that full time, evaluating model performance and finding applications for the software.

So, how good are AlphaFold's predictions?

In 2020 I compared AlphaFold's predictions to the structures found in experimental studies reported in the highest-impact journals, mostly those published in *Nature*. At the time we were trying to predict single-chain protein structures, and AlphaFold did really rather well. But I noticed that many of the papers weren't looking at single chains, they were studying more complex systems that contained multiple chains.

That motivated us to start working on AlphaFold Multimer, a version of the model specifically trained for multichain protein complexes.

Have AlphaFold's predictions ever disagreed with experimentally derived structures, which were then found to be wrong?

There have been a few cases; but they weren't ones that I found. Since AlphaFold became available for anyone to use, researchers have carried out an enormous number of investigations with the software. One finding that came out of that effort is, in some instances, AlphaFold predicts more accurate structures than have been experimentally found with nuclear magnetic resonance (NMR) techniques. In NMR, the experimental data need quite a lot of processing to turn them into a structure. And there have been instances where AlphaFold's predicted structure has fit the data better than the original NMR-derived one.

How many structures has AlphaFold predicted to date?

Over 200 million.

Any notable proteins whose structures you have worked on?

With the version of AlphaFold evaluated in CASP14 (the 14th iteration of a biennial assessment of protein-structure-prediction models), the first sequence I worked on was for one of the proteins of SARS-CoV-2, the virus that causes COVID-19. That was a sad way to start testing the system, but people were obviously interested in what that protein's structure looked like.

What's on the horizon for AlphaFold?

I can't share many details, but I can say that the team behind AlphaFold is committed to working on protein-related problems for the long-term. There are still lots of things AlphaFold can't do, such as modeling the nonprotein components bound to the system of interest or the influence of water molecules or ligands on how a given protein behaves. The 3D structure of a protein is also just one of its properties. It would be cool to be able to predict other things, such as how a protein's shape is affected by point mutations.

There are about 20 people working on updates to AlphaFold—its success is really a team effort—and the team is constantly collaborating with researchers to make sure we are looking at problems that are of interest to scientists. We have a constant stream of follow-up problems to investigate.

Katherine Wright is the Deputy Editor of *Physics Magazine*.