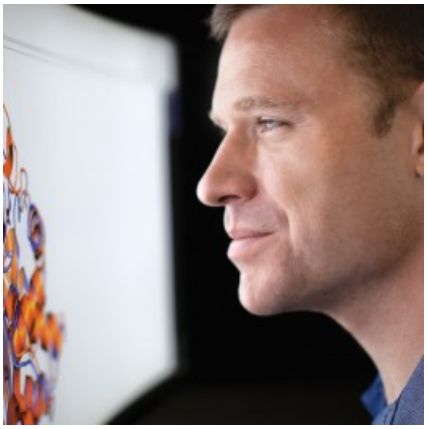
COVID Information Commons (CIC) Research Lightning Talk

Transcript of a Presentation by Gregory Bowman (Washington University in St. Louis), February 10, 2021



Title: RAPID: Folding@home and COVID-19

Gregory R Bowman CIC Database Profile

NSF Award #: 2032663

YouTube Recording with Slides

February 2021 CIC Webinar Information

Transcript Editor: Macy Moujabber

Transcript

Slide 1:

Thanks, appreciate you all being here. So, my group and I are really obsessed with protein dynamics, and just to give a little bit of background make sure we're all on the same page. Proteins are these molecular machines that are responsible for many of the active processes that we associate with life, so everything from muscle contraction to the detection of light in the eye, and to perform all these functions like machines we work with on microscopic scales that have many moving parts.

Slide 2:

But experimentally often all that we can see are these static snapshots of what a protein typically looks like. So, this is an atomically detailed structure of one of our favorite proteins. So, the spheres represent all the atoms in this protein and this is a really rich source of information so we can immediately see what's called the active site so this pocket where this particular protein catalyzes a chemical reaction. But again, we know that there's a lot more to the story so as rich as the information content of this one structure is, it's just the tip of the iceberg. And so, one of our specialties is using computer simulations to simulate how every atom protein like this moves over time and improve our understanding of how these machines function and how we can control them. So, for example here we can see the opening of an unexpected or cryptic pocket, as we call it, that we've since shown is implicated in the function of this protein and presents new opportunities for drug design for example. Alright, so these are really powerful approaches and we've made a lot of success in making quantitative connections with experiments, but they're also extremely computationally demanding. So, simulating the sorts of time scales we would like to capture could easily take hundreds to tens of millions of years depending on the protein in question if we were to use a single powerful computer.

Slide 3:

And so, one of the things that we in a half dozen other research labs around the world do is run a distributed computing project called Folding at Home, where we ask anyone with a computer and an internet connection to volunteer to help us run simulations on their personal computers and send us back the data. So, this is a map showing a pin prick of light where everyone has been participating in Folding at Home. And at the beginning of 2020 we had about 30,000 devices around the world actively participating in this project, helping us to run calculations that probably would have cost millions of dollars on the cloud by any other means. And so, this was an extremely powerful tool and as it became clear that COVID-19 was going to become a pandemic, we realized that we had the opportunity to bring this amazing tool to bear to understand all the protein components of the virus and identify new therapeutic opportunities and maybe inform the development of vaccines. And so, at the end of February of this past year we launched our first simulations of SARS-CoV-2 proteins, and the response globally was simply amazing.

Slide 4:

So, I remember I said that we had 30,000 devices participating at the beginning of 2020. Here in blue I'm showing the cumulative downloads of our software over the first few months of the pandemic and you can see that within the first couple of months we rose to having well over a million devices around the world actively participating and Folding at Home at a given time. And so, the upshot of this is that Folding at Home became the most powerful supercomputer in the world the first to measure its performance in units called exaflox [?] and suffice it to say that we had five times very conservatively estimate we had five times the raw computing power of the world's fastest supercomputer at the time which is the Summit supercomputer in the U.S. So, with all this compute power we set to work simulating every protein we could from the viral proteome as well as other coronaviruses and human proteins that are involved in the immune response or activity of the proteins from the virus.

Slide 5:

I don't have time obviously to go into all of those, but one of my favorites is our work on the spike. So, as you probably know going to all these red protrusions and this canonical image of the virus are called spikes. Each is actually a complex of three proteins. And what really fascinated us is that when you see this image, the structure of the spike you see is actually what we call the closed states these three proteins are curled up on each other and they're actually burying the interface that is responsible for binding to a protein called ACE2 on a potential host and initiating infection. And so, as you can imagine that can't happen in this closed-up state that we see and so somehow these proteins have to open up like the mouth of one of these Demogorgon monsters from the television series *Stranger Things* in order to expose this key binding interface to latch on to a potential host. And so, the reason this the virus has this opening and closing is because in the closed state the virus can hide from being detected by our immune system better than if it was open all the time, exposing this sensitive site.

Slide 6:

So here I'll just show you a bit of the simulation. So, the three colors here are the three different proteins that make up one spike. The top left here is tethered to the surface of the virus and the bottom right is where the spike has to open up in order to bind to a human cell, for example. And if we were to tackle this with a conventional simulation from our perspective, our community's perspective, this is a huge system. 3,600 amino acid residues that make up these proteins and so all we would see is wiggling around this closed state but with the power of Folding at Home, we're actually able to see this dramatic structural change unfold over time where one of the three protein components opens up quite widely and exposes this surface here that was buried in the closed state and again is responsible for binding to ACE2 and initiating infection. So, this is really cool because it explains how that interface gets exposed. It also predicts that there's all kinds of other surfaces that actually get exposed by this opening motion and those are exciting because many could be epitopes for antibodies or the targets of small molecule drugs. And now we have structural information on what that looks like we could use to inform the development of vaccines, and we're also building models for all of the emergent variants and trying to understand how they might change the spikes behavior and yes, the efficacy of vaccines, for example, and inform the design of new vaccines. As I said, I've been doing the same sorts of things with every other protein from the virus we can. All the data is available on AWS and OSF if you're interested in playing with it and you can see our bio archive paper for the links.

Slide 7:

Now with that, I'm extremely grateful for the Folding at Home team. Both scientists all spread around the globe and also our volunteers who are even more globally distributed and all of our funding sources including the, you know, NSF for their help keeping this going as we struggled at the beginning to scale up in response to the massive increase in participation in this project. Thanks.