Feature



CORONAVIRUS PIECE BY PIECE

Biologists are working at breakneck speed to solve the structures of key SARS-CoV-2 proteins and use them against the virus. **By Megan Scudellari**

ying in bed on the night of 10 January, scrolling through news on his smartphone, Andrew Mesecar got an alert. He sat up. It was here. The complete genome of a coronavirus causing a cluster of pneumonia-like cases in Wuhan, China, had just been posted online.

Around the world, similar notifications appeared on the devices of scientists who first crossed swords with coronaviruses in the 2003 outbreak of SARS (severe acute respiratory syndrome) and then again with MERS (Middle East respiratory syndrome) in 2012. Instantly, the researchers mobilized against a new adversary. "We always knew that this was going to come back," says Mesecar, head of biochemistry at Purdue University in West Lafayette, Indiana. "It's what history has shown us."

In Lübeck, Germany, Rolf Hilgenfeld stopped packing boxes for his retirement and started preparing buffers for crystallography. In Minnesota, Fang Li stayed up all night analysing the new genome and drafting a manuscript. In Shanghai, China, Haitao Yang rallied a dozen graduate students to clear their schedules. In Texas, Jason McLellan instructed laboratory members to start assembling gene sequences from the viral genome.

Within 24 hours, a network of structural biologists around the world had redirected their labs towards a single goal – solving the protein structures of a deadly, rapidly spreading new contagion. To do so, they would need to sift through the 29,811 RNA bases in the

virus's genome, seeking out the instructions for each of its estimated 25–29 proteins. With those instructions in hand, the scientists could recreate the proteins in the lab, visualize them and then, hopefully, identify drug compounds to block them or develop vaccines to incite the immune system against them.

"Here we go," thought Mesecar. "I'd better get some sleep."

11 January: 41 confirmed cases of COVID-19 worldwide

Mesecar woke at 6 a.m. the next day, turned on the coffee pot and began blasting through the new genome looking for recognizable protein sequences. It didn't take long. He had spent 17 years studying coronaviruses, and the new virus's genome looked very familiar. "Holy shit," he thought. "This is the same thing as SARS."

Right away, Mesecar contacted Karla Satchell, a microbiologist at Northwestern University in Chicago, Illinois. Satchell is co-director of the Center for Structural Genomics of Infectious Diseases (CSGID), a consortium of eight institutions set up exactly for moments like this – to rapidly investigate the structures of emerging infectious agents.

To solve the 3D structure of a protein at high resolution, scientists first design a gene construct - a circle of DNA containing the instructions for the protein, together with regulatory sequences to control where and how it is expressed. They then insert the construct into living cells, often the bacterium Escherichia coli, using the cells' own machinery to churn out the desired protein. Next, they purify the protein so that they can visualize its structure using either of two methods. One is X-ray crystallography, which involves growing tiny crystals of pure protein and revealing their internal structure by bombarding them with X-rays from a high-energy electron beam. The other is cryo-electron microscopy (cryo-EM), a process of scanning flash-frozen proteins using a high-powered electron microscope.

Either process can take months, even years, for an unfamiliar protein. Luckily, many of the new coronavirus proteins were familiar, with 70–80% sequence similarity to SARS-CoV, the virus that caused the 2003 SARS outbreak. By 7:30 a.m., Mesecar and his team had begun designing gene constructs for the new viral proteins, and even predicted which of their existing coronavirus inhibitors might block these proteins.

Satchell, who had been following early news reports about the virus, organized a virtual meeting of consortium members to start solving the virus's proteins. "We've thrown the weight of every investigator at every site behind COVID," says Satchell. Mesecar, a CSGID investigator, started with M^{pro}, the virus's main protease, an enzyme that cuts out proteins from a long strand that the virus produces when it invades a cell, like a tailor cutting out pattern pieces. Without M^{pro}, there is no viral replication. Humans do not have a similar protease, so drugs targeting this protein are less likely to cause side effects.

13 January: 42 confirmed cases

In McLellan's molecular biosciences lab at the University of Texas at Austin, graduate student Daniel Wrapp spent the weekend designing a gene construct for another key protein – the outer, three-pronged spike that gives the coronavirus its crown-like appearance and name. Wrapp placed an order for the constructs with a commercial firm that Monday, 13 January.

McLellan had been involved in determining

THE KEY CORONAVIRUS PROTEINS

Researchers are racing to visualize and understand the proteins used by SARS-CoV-2 to enter cells and replicate. That information could be crucial for making drugs and vaccines to stop the virus.



the structures of two other coronavirus spikes - from HKU1, a cause of common colds¹, and from the MERS virus². The work was done in collaboration with structural biologist Andrew Ward at the Scripps Research Institute in La Jolla, California, and virologist Barney Graham at the US National Institute of Allergy and Infectious Diseases' Vaccine Research Center in Bethesda, Maryland. So, the group knew how to tweak the spike protein's genetic sequence so that it would stabilize in a pre-fusion shape - the form it adopts before it docks onto a host cell. "Our ability to get this particular structure was based upon all our prior knowledge from working on HKU1 and MERS and SARS," says McLellan.

While McLellan's team waited for the construct to arrive, Graham called Moderna Therapeutics, a drug-discovery company in Cambridge, Massachusetts, with which the Vaccine Research Center had been working on a pandemic-preparedness project. On 13 January – before any spike protein had been made – Moderna began preparing its manufacturing facilities to make a coronavirus vaccine based on that protein.

26 January: 2,014 confirmed cases

At ShanghaiTech University in China, Zihe Rao, Haitao Yang and their colleagues worked day and night, sacrificing their week-long Chinese Lunar New Year holiday, to solve the M^{pro} structure and those of another trio of proteins that the coronavirus uses to replicate.

Using X-ray data acquired at the Shanghai Synchrotron Radiation Facility and the National Center for Protein Science Shanghai – which both allocated special beam time for the project – the team solved the crystal structure of M^{pro} bound to an inhibitor³. In 2003, it had taken them two months to solve the structure of the SARS-CoV main protease. This time, it took one week. M^{pro} in coronaviruses is made up of two identical subunits and looks like a motheaten heart, with an active enzyme site on each side of the structure. On 26 January, Rao and Yang submitted the M^{pro} structural data to the Protein Data Bank (PDB), an open-access digital resource for 3D structures of biological molecules. By 5 February, the data had been processed and the final structure was released online – not a moment too soon, says Yang. The laboratory had already received an overwhelming 300 requests for the structure.

While working on M^{pro}, Rao contacted a former co-worker, David Stuart, a structural biologist at the University of Oxford, UK, who is life-sciences director at Diamond Light Source, the United Kingdom's synchrotron facility. The UK and Shanghai groups began collaborating closely to share advice and avoid overlap, says Martin Walsh, deputy lifesciences director at Diamond. "We keep each other up to date on things, and try to benefit from the different approaches they're using and we're using."

Because the Shanghai team solved M^{pro} in complex with an inhibitor, the Diamond team decided to focus on crystallizing the protein with no molecule attached, hoping to identify active sites to which potential drug compounds might bind. Over two weeks, Walsh's group ran 17,000 experiments to hit on the best recipe for precipitating the unbound protein into a crystal.

1 February: 11,953 confirmed cases

In Hilgenfeld's lab at the University of Lübeck, researcher Linlin Zhang had taken to phoning the company making the M^{pro} gene construct daily until it finally arrived. Thanks to the lab's experience crystallizing other coronavirus proteases, Zhang grew M^{pro} crystals in 10 days, and on 1 February, she took the precious samples to the BESSY II synchrotron in Berlin, which

Feature

THE SPIKE LOCKS ON

Binding regions on the tip of the spike open out to attach to the human ACE2 receptor, found on lung cells and elsewhere in the body.



opened up a beamline especially for the project. In addition to focusing on the unbound M^{pro} structure, Hilgenfeld docked a small-molecule inhibitor called 13a, which he had designed to inhibit the MERS virus, into the protein's active site. It wasn't a perfect fit, so the team altered a residue on the compound and named it 13b. This one "fit nicely", says Hilgenfeld, and in ten more days his team had solved the structure

of M^{pro} bound to the inhibitor⁴. McLellan's group in Texas was solving the spike protein structure at similar speed. As soon as the group had finished gathering high-resolution electron-microscopy data of the stabilized spike, thanks to a multimillion-dollar cryo-EM facility at the university. McLellan sent the data to Graham at the Vaccine Research Center.

Vaccines are often based on presenting parts of a virus to the human immune system to provoke a response, and the spike protein is an obvious candidate because it has a crucial role in infection.

The spike is formed of three identical molecules stuck together in the shape of a pyramid, with a hinge-like trapdoor. This opens to expose a portion that grabs onto a receptor on a human cell (see 'The key coronavirus proteins'). Graham and McLellan's past work on a similar protein⁵ suggested that presenting the spike protein in its pre-grab state would provoke the human immune system. From the complete structure, Graham could see that McLellan's gene construct made a high-quality protein arranged in the right conformation. "It was really, really important to have that electron-microscopy information," says Graham.

Graham tested the spike protein in mice, working to improve its expression levels and the strength of its effect on the immune system, and sent the sequence to Moderna, where the production line was ready and waiting. On 7 February, Moderna completed its first batch of the vaccine based on that protein.

Spike protein

Meanwhile.on 10 February.just 12 days after harvesting the protein, McLellan and his group submitted its cryo-EM structure⁶ to the PDB. By studying the spike in detail, they found that it binds to its human cell receptor, a protein called ACE2, at least ten times more tightly than SARS-CoV does.

At the University of Minnesota in Saint Paul, Li's team was on its way to working out why. On 11 February, Li and his colleagues began collecting X-ray data from the spike protein using the Advanced Photon Source (APS), the synchrotron facility at the US Department of Energy's Argonne National Laboratory near Chicago, Illinois. By 13 February, the researchers had defined the small, important spot where the spike protein locks on to the ACE2 receptor⁷ (see 'The spike locks on'). They found that the new coronavirus spike protein has small molecular differences in its binding region compared with that of SARS-CoV, which might be why the new virus attaches to ACE2 more strongly. These changes could also explain why it seems to infect cells better and spreads faster than the SARS virus. That same week, the virus also got a name: SARS-CoV-2.

18 February: 73,332 confirmed cases

By mid-February, protein structures were pouring out. On 18 February, Hilgenfeld, Zhang and their colleagues submitted a paper⁴ on the M^{pro} structure alone and bound to 13b, and posted the preprint on the bioRxiv server on 20 February. "It was pretty fast," Hilgenfeld admits. "The longest time period was just getting it published." That same day, the Diamond team released the high-resolution

crystal structure of unbound Mpro on its website (go.nature.com/2z41uwi).

To support US teams, the APS and other national synchrotrons coordinated their schedules to ensure there would be no interruption in beamtime if one facility had to close for maintenance or because of a local outbreak. "Our goal is just to keep the research going," says Stephen Streiffer, director of the APS. "The rate at which people are working at this is an order of magnitude faster than they've been able to work on other problems."

So far, the CSGID consortium has solved 12 unique SARS-CoV-2 protein structures. which are kept in a new online database with their accompanying genomic information (see https://coronavirus3d.org). "We've been part of projects like this on cancer, but it took five years to set that all up," says Adam Godzik, a bioinformatician at the University of California, Riverside, and a CSGID investigator. "This happened spontaneously in the course of months."

16 March: 167.515 confirmed cases

With 3D structures in hand, structural-biology teams moved straight to next steps. "Structures aren't everything," says Mesecar. "You want to get to compounds - to antivirals and vaccines."

On 16 March, just 65 days after the viral genome was released, clinicians gave the first dose of Moderna's vaccine candidate to a patient in a clinical trial funded by the US National Institutes of Health.

"It was a lot faster than even the fastest one we'd previously done," says Graham. Because of research on SARS and MERS, coronaviruses were probably the only viral family for which that was possible, he adds. "If it was a bunvavirus or an arenavirus, we would have been lost for two to three years."

But even a vaccine developed at record-breaking speed is likely to be a slower solution than repurposing an approved drug, or at least finding one for which safety testing has begun. "That's absolutely going to be the fastest way to help patients sick in the hospital today," says Satchell.

That was exactly what Andrew Hopkins was planning. On 19 March, Hopkins, the chief executive of Exscientia, an artificial-intelligence drug-discovery company in Oxford, UK, took delivery of a large styrofoam cooler packed with dry ice. Inside was a library of 12,000 drug compounds known to be safe and ready for human use, sent from Scripps Research in California. The Exscientia team, working closely with Diamond, immediately began screening the collection against four of Diamond's structures: M^{pro}, the spike protein, a second protease and the replication-machinery complex. Exscientia is currently preparing to test compounds that bind to the first two proteins for antiviral activity, says Hopkins.



Once the virus has fused with a host cell, the virus



Similarly, the ShanghaiTech team conducted virtual and high-throughput screening of a library of more than 10,000 approved drugs and compounds already in clinical trials, to see whether any would disable M^{pro}. They identified six promising candidates3. One of them, ebselen, is already in clinical trials for the treatment of bipolar disorder and hearing loss, and the group is preparing animal tests to study its activity in vivo, says Yang.

On 10 April, Rao, Yang and their collaborators published⁸ the structure of the virus's replication complex - a large protein called RNA-dependent RNA polymerase (RdRp, or nsp12) that forms a complex with two others, nsp7 and nsp8. They also modelled how it binds to the antiviral drug remdesivir, originally developed to treat Ebola and now in phase III trials for coronavirus. Another recently completed structure of the protein in complex with the drug⁹ could provide a template to help model and modify other existing antivirals.

22 April: 2,471,136 confirmed cases

The hard-core biochemistry of designing brand-new, custom drugs to inhibit SARS-CoV-2 proteins will take months, even years, but could eventually lead to the best-performing drugs against the infection (see 'Breaking the cycle').

The ShanghaiTech team and collaborators have designed and synthesized a series of compounds targeting the active site of Mpro. On 22 April, after much chemical tweaking, they published details of one that inhibits viral replication in cells and was not toxic when tested in rats and dogs¹⁰. The team will continue developing that compound as a drug candidate, says Yang.

The Diamond team has identified

91 chemical fragments - bits of molecules that are less than one-third the size of a normal drug - that bind to M^{pro}. Those fragments inspired the launch of a non-profit crowdsourced initiative, the COVID Moonshot, to engage chemists around the world to use the fragments to design antiviral drug candidates. The initiative has received more than 4,600 design submissions, and several therapeutic possibilities are already emerging.

In Germany, researcher Katharina Rox at the Helmholtz Centre for Infection Research in Braunschweig tested Hilgenfeld's 13b

"Structures aren't everything. You want to get to compounds - to antivirals and vaccines."

compound in mice, showing that it was safe and accumulated well in the lungs⁴, a key infection site. Meanwhile, a compound that Mesecar developed to inhibit SARS-CoV, compound 77, has been shown in unpublished work to have antiviral activity against SARS-CoV-2 in cells, and he hopes to complete animal studies by the end of the summer.

14 May: 4,248,389 confirmed cases

Structural biologists continue to plug away at the remaining unsolved proteins in the coronavirus genome. These include ORF8, a protein whose function remains mysterious. "We predict it should be crystallizable, but nobody has done it, so we're trying," says Godzik.

In the United Kingdom, the Diamond team is screening various compounds against a second coronavirus protease. In Texas, McLellan has shipped spike constructs to more than 100 labs worldwide. Many are looking for treatments, using the protein to fish antibodies out of the blood of people who have had COVID-19, and McLellan's team is now characterizing the first of these potentially therapeutic antibodies.

Hilgenfeld, who was officially scheduled to retire on 1 April as a result of a mandatory retirement policy, has packed up his office but continues to work. "I've been working on coronaviruses for 20 years, and most of the time it was neglected and not taken seriously," he says. "Now that it's happened, how can I leave?" His team is investigating other SARS-CoV-2 structures, including nsp3, a large protein that the virus uses to shut down host-cell defences.

The race against the virus can't afford to slow down anytime soon. As soon as countries start lifting restrictions on people's movement, the virus will return and "flip around the world again", says Satchell. "When that happens, it would be really great to have beautiful drugs that were designed specifically to target this coronavirus," she says. "But we need to do it fast."

Megan Scudellari is a science journalist in Boston, Massachusetts.

- 1. Kirchdoerfer, R. N. et al. Nature 531, 118-121 (2016).
- Pallesen, J. et al. Proc. Natl Acad. Sci. USA 114, E7348-2.
- E7357 (2017). 3. Jin, Z. et al. Nature https://doi.org/10.1038/s41586-020-2223-y (2020).
- Zhang, L. et al. Science 368, 409-412 (2020).
- McLellan, J. S. et al. Science **342**, 592–598 (2013). 5
- 6. Wrapp, D. et al. Science 367, 1260-1263 (2020).
- Shang, J. et al. Nature 581, 221-224 (2020). 7.
- 8. Gao, Y. et al. Science 368, 779-782 (2020) Yin, W. et al. Science https://doi.org/10.1126/science.
- 9. abc1560 (2020)
- 10. Dai, W. et al. Science https://doi.org/10.1126/science. abb4489 (2020).