The proteins of Covid-19: Machine learning assisted structure analysis reveals SARS-CoV-2 virus tactics

The proteins of SARS-CoV-2 play key roles in how the virus manages to evade immune defense and replicate itself in patients' cells. An international research team – with significant contribution from the Technical University of Munich (TUM) – has now compiled the most detailed view of the virus' protein structures available to date. The analysis employing artificial intelligence methods has revealed surprising findings.

How does the SARS-CoV-2 virus manage to evade immune defense and replicate itself in patients suffering from COVID-19? To shed light on this question, an international research team has assembled the most comprehensive view to date on the precise, three-dimensional shape of each SARS-CoV-2 protein – including the well-known spike protein.

To assemble this map, the team used high-throughput machine learning, an approach that enabled them to predict structural states likely to occur in coronavirus proteins, based on states that have been seen in related proteins. The final map consists of 2,060 atomic-resolution 3D models involving the coronavirus's 27 proteins. All structural models are freely available through the Aquaria-COVID resource (https://aquaria.ws/covid).

"This provides an unprecedented wealth of detail that will help researchers better understand the molecular mechanisms that cause COVID-19, and will help in developing therapies to fight the pandemic, for example, by identifying potential new targets for future treatments or vaccines," says Burkhard Rost, professor of bioinformatics at the Technical University of Munich.

The structural coverage map – key to a wealth of knowledge

A second part of the study used a complementary approach, known as human-in-the-loop machine learning, to create a novel, visual interface that summarizes everything currently known – and not known – about the shape of SARS-CoV-2 proteins.

The visual interface, called a structural coverage map, can also be used as a navigation aid for researchers, helping them find and use structural models that relate to specific research questions. And these models have already revealed several vital clues into how coronaviruses are able to take over our cells.

How coronaviruses take over our cells

Using their machine learning approaches, the team was able to identify three coronavirus proteins (NSP3, NSP13, and NSP16) that 'mimic' human proteins, successfully fooling a host cell into believing that they are native proteins operating in the best interest of the cell.

The modelling also revealed five coronavirus proteins (NSP1, NSP3, spike glycoprotein, envelope protein, and ORF9b protein) that the researchers say 'hijack' or disrupt processes in human cells, thereby helping the virus take control, complete its life cycle and spread to other cells.

Understanding how the virus operates – and how to stop it

"In analyzing these structural models, we also found new clues into precisely how the virus copies its own genome – which is the central process that allows the virus to spread rapidly in any person unlucky enough to become infected," says Burkhard Rost. "The insights from our study take us closer to understanding how the virus operates, and what we can do to stop it."

"The longer the virus circulates, the more chances it has to mutate and form new variants such as the Delta strain," says Professor O'Donoghue, the lead author of the study from the Garvan Institute in Sydney. "Our resource will help researchers understand how new strains of the virus differ from each other – a piece of the puzzle that we hope will help in dealing with new variants as they emerge."

###

The team includes researchers of the Garvan Institute of Medical Research, the Commonwealth Scientific and Industrial Research Organisation (CSIRO) and the University of New South Wales (Sydney, Australia), the Technical University of Munich (Garching/Munich, Germany), the University of Applied Sciences Weihenstephan-Triesdorf (Freising, Germany), the University of Dundee (Scotland) and the University College London (UCL, London, UK).

wissenschaftliche Ansprechpartner:

Prof. Dr. Burkhard Rost
Technical University of Munich
Professorship of Bioinformatics
Boltzmannstr. 3, 85748 Garching, Germany
Tel.: +49 89 289 17811 (Office)
E-mail: assistant@rostlab.org

Originalpublikation:

Sean I. O'Donoghue, Andrea Schafferhans, Neblina Sikta, Christian Stolte, Sandeep Kaur, Bosco K. Ho, Stuart Anderson, James B. Procter, Christian Dallago, Nicola Bordin, Matt Adcock, Burkhard Rost
SARS-CoV-2 structural coverage map reveals viral protein assembly, mimicry, and hijacking mechanisms
Molecular Systems Biology, Sept. 14, 2021 – DOI: 10.15252/msb.202010079
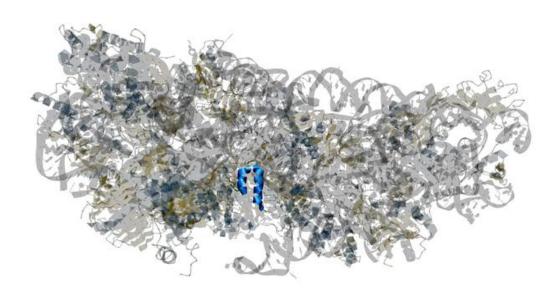
URL zur Pressemitteilung: https://www.embopress.org/doi/10.15252/msb.202010079 Original publication

URL zur Pressemitteilung: https://www.tum.de/en/about-tum/news/press-releases/details/36913 Press release on the TUM website

URL zur Pressemitteilung: https://www.rostlab.org/ Website of Prof. Rost's research group

URL zur Pressemitteilung: https://aquaria.ws/covid SARS-CoV-2 protein structure database

Prof. Dr. Burkhard Rost, Professorship of Bioinformatics, Technical University of Munich
Juli Eberle / ediundsepp / TUM

Structure of the SARS-CoV-2 protein NSP1 (blue) in complex with a host ribosome (grey). NSP1 blocks the translation of host mRNA in infected cells.
Seán O'Donoghue
Garvan Institute of Medical Research