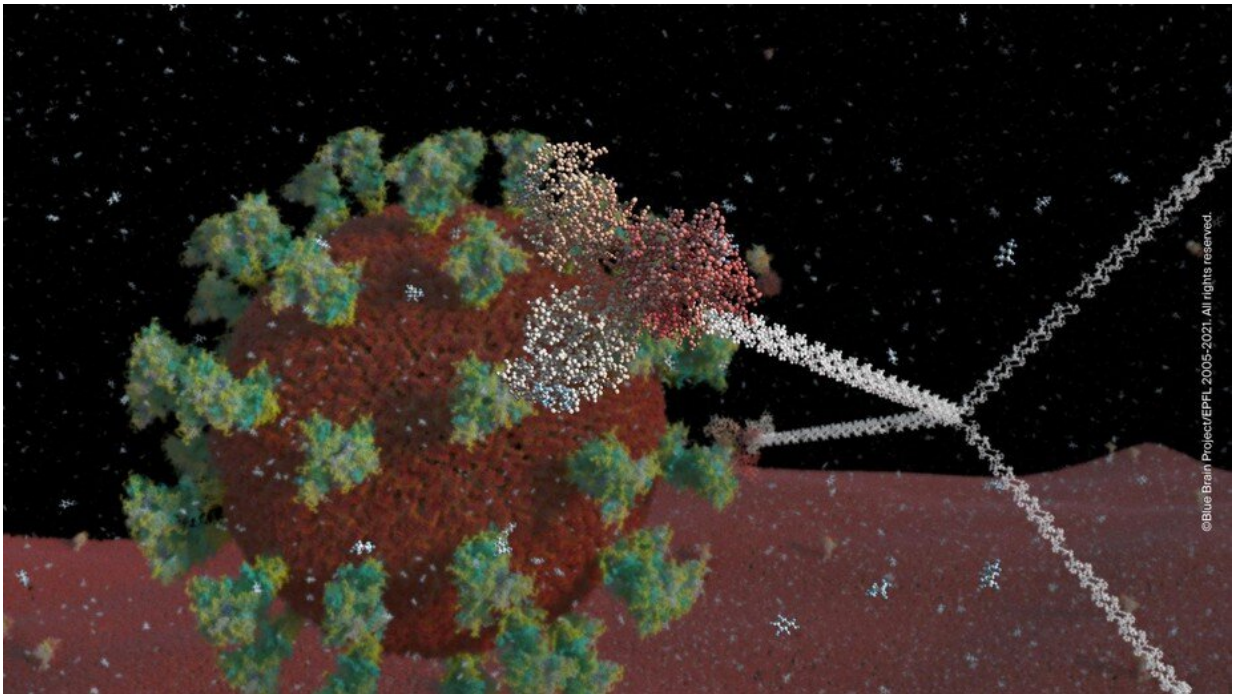# AI reveals how glucose helps the SARS-CoV-2 virus

August 2 2021

Digital reconstruction of SARS-CoV-2 virus in the lung environment. Credit: Blue Brain Project

Why do some people get sick and die from COVID-19 while others seem to be completely unaffected? EPFL's Blue Brain Project deployed its powerful brain simulation technology and expertise in cellular and molecular biology to try and answer this question.

A group in the Blue Brain assembled an AI tool that could read hundreds of thousands of [scientific papers](link), extract the knowledge and assemble the answer—"A machine-generated view of the role of blood glucose levels in the severity of COVID-19" was published today by *Frontiers in Public Health*

In response to the COVID-19 pandemic, the COVID-19 Open Research Dataset (CORD-19) of over 400,000 scholarly articles was made [open access](link), including over 150,000 with full text papers related to COVID-19, SARS-CoV-2, and other coronaviruses. The CORD-19 dataset is the most extensive coronavirus literature collection available for [data mining](link) to date and the coalition behind it has challenged AI experts to apply their skills in natural language processing and other machine learning techniques in order to generate new insights that may help in the ongoing fight against COVID-19.

"Since early 2020, Blue Brain has been proactively contributing to the fight against COVID-19," explains Prof. Henry Markram, Founder and Director of the Blue Brain Project. "With this call to action, we realized we could use our Machine Learning technologies and Data and Knowledge Engineering expertise to develop text and data mining tools required to try and help the medical community. Blue Brain set out to answer one of the most puzzling aspects of this pandemic—why some people get very sick, while others are completely unaffected."

Building and using the text and data mining tools

Accordingly, Blue Brain built and trained machine-learning models to mine these papers and extract structured information from text sources. A simple analysis by this text mining toolbox "Blue Brain Search' of the CORD-19v47 dataset revealed papers that all pointed to [glucose metabolism](link) as the most frequently mentioned biological variable.

Using [Blue Graph](link), a unifying Python framework that analyzes extracted text concepts to construct knowledge graphs, the group constructed specific knowledge graphs to focus on all the findings that considered glucose in the context of respiratory diseases, coronaviruses, and COVID-19. This allowed for the exploration of the potential role of glucose across many levels, from the most superficial symptomatic associations to the deepest biochemical mechanisms implicated in the disease.

From the facts and findings of thousands of papers mined, multiple lines of evidence emerged that elevated blood glucose levels were either caused by abnormal glucose metabolism, or induced during hospitalization, drug treatments or by IV administration. This approach correlated extremely well with COVID-19 severity across the population and revealed how elevated glucose helps virtually every step of the viral infection, from its onset in the lungs, through to severe complications such as Acute Respiratory Distress Syndrome, multi-organ failure and thrombotic events.

"Subsequently, in the paper, we discuss the potential consequences of this hypothesis and propose areas for further investigation into diagnostics, treatments and interventions that may help to reduce the severity of COVID-19 and help manage the public health impact of the pandemic," discloses Blue Brain's Molecular Biologist Dr. Emmanuelle Logette.

The potential of open access scientific papers

"Scientists immediately went to work when the pandemic started and within a year published over a hundred thousand papers. But, can anyone read so many papers? Can anyone see and understand all the patterns across all this research?" asks Prof. Henry Markram. "Fortunately, the coalition behind the CORD-19 dataset convinced all subscription

publishers to bring these papers over the subscription paywall and make them openly accessible so that they can be mined with modern machine learning and knowledge engineering technologies."

"With access to the CORD-19 dataset, Blue Brain quickly assembled an AI tool and targeted it to try and find out why some get sick and others not. Is it enough to just say that older people are more vulnerable? We must find out why. Why do some apparently healthy people die from COVID-19? Why do so many people die in the ICU? To answer these questions, we directed our AI to trace every step of the viral infection from the moment the virus enters the lungs until the time when the virus breaks out of the cells in the lungs and spreads throughout the body to infect the organs," explains Prof. Markram. "We also built the virus at an atomistic level and developed a computational model of the infection so we could try to test what was coming out of the literature. I think we did find the most likely reason why some people get sicker than others," he concludes.

An example of this is the team using [Blue Brain BioExplorer](link) to visually show the main impacts of high glucose in airway surface liquid on the primary step of infections in the lung and explaining the increased susceptibility to respiratory viruses in at-risk patients.

Blue Brain BioExplorer was built to reconstruct, visualize, explore and describe in detail the structure and function of the coronavirus for this study, and is open source for others to use to answer key scientific questions.

"Pioneering Simulation Neuroscience to better understand the brain has numerous collateral benefits," states Prof. Markram. "This study shows how Blue Brain's computing technologies and unique team of multi-disciplinary experts can quickly be redirected to help in a global health crisis."

A major step forward for science and understanding the brain

"The COVID-19 study also shows why we believe that computational tools are so important to help us understand the brain," explains Prof. Markram. "The problem is even bigger. There are several million scientific papers that one would need to read and understand to work out what we know about the brain. Does anyone know what we know? But, machines can read so many papers. This is the reason that the Blue Brain has developed some of the most advanced knowledge engineering, mathematical and machine learning accelerator technologies. Actually, this solves only a part of the challenge. With an AI tool that can read all these papers, we would still only know only a small fraction of what the [brain](#) contains and how it works. But building model brains using design principles, helps us to try and complete the picture." he concludes.

Is it right to only open science during a pandemic?

Prof. Markram also expressed his frustration with the all too common practice of locking up of scientific knowledge by subscription publishers. "When the CORD-19 literature dataset was made available to us, we at Blue Brain were able to point our technology at COVID-19 and propose an answer to an important question in the battle against this deadly virus. Therefore, is it right to only make science papers (that are publicly funded) open to the public during a pandemic when the same kind of techniques can be used to help address so many other diseases, accelerate science, and help save the planet from climate change?"