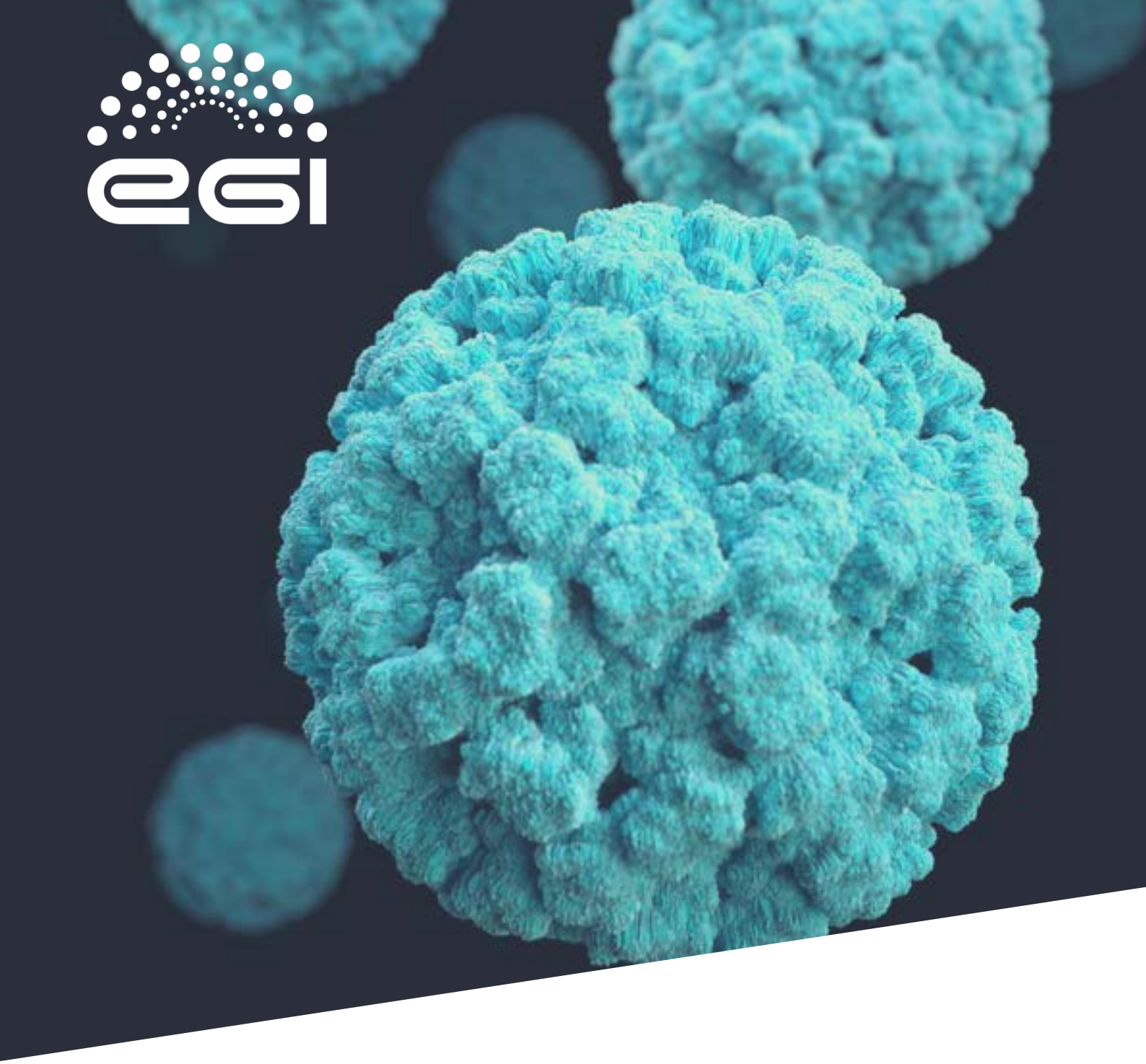


EGI SUPPORTING COVID-19 RESEARCH

E-Infrastructure

E-INFRASTRUCTURES PROVIDE DIGITAL-BASED SERVICES AND TOOLS FOR DATA- AND COMPUTING-INTENSIVE RESEARCH IN VIRTUAL AND COLLABORATIVE ENVIRONMENTS.

What is EGI?

THE EGI FEDERATION AND ITS IMPORTANCE

The EGI Federation is an international e-Infrastructure set up to provide advanced computing and data analytics services for research and innovation. The EGI e-infrastructure is publicly-funded and comprises hundreds of data centres and cloud providers spread across Europe and worldwide.

The EGI Federation offers a wide range of services for compute, storage, data and support, and provides access to **+1,211,233** computing cores and **+1,190 PB** of disk and tape storage.

Vision

All researchers should have seamless access to services, resources and expertise to collaborate and conduct world-class research and innovation.

Mission

Deliver open solutions for advanced computing and data analytics in research and innovation.

Advanced Computing for Research

1. EGI fast-track call for COVID-19 research projects

The EGI Federation and the Open Science Grid are calling for projects addressing COVID-19 scientific questions that can benefit from technical enabling support and a large scale compute and storage capacity to accelerate their data analysis tasks, leveraging the capabilities of the EGI Federation, a network of more than 250 data centres worldwide serving data-intensive science.

2. WeNMR and the HADDOCK portal

HADDOCK is a WeNMR portal for the modeling of biomolecular complexes. It can support complex simulation projects thanks to the access to large scale research compute facilities in the world. In the context of COVID-19 related research it can, for example, model interaction between virus proteins and human one, or support the docking of small molecules to targets such as the COVID-19 protease. The tool offered by the Bijvoet Center of the Utrecht University, Netherlands, is supported by the EOSC-hub H2020 project and is accessible through the European Open Science Cloud (EOSC) Portal.

3. EOSC-Synergy Galaxy Service

The EOSC-Synergy project is providing a Galaxy portal freely accessible to users and provides (1) data on coronavirus and specially samples from SARS-CoV-2 (COVID-19) daily updated from the public international databanks, and (2) some key tools for identification of mutations, phylogenetic analysis, sample processing and visualization. The portal is an open lab for researchers that want to run their experiments without the burden of downloading data and installing tools, with easy tools for sharing data, workflows and results.

Collaboration with OPEN SCIENCE GRID

Open Science Grid

The Open Science Grid (OSG) facilitates access to distributed high throughput computing for US based researchers and their international collaborators. Researchers can submit batch jobs from their home institutions, or receive accounts on OSG supported submit points, to run simulations or analyze data across more than 100 clusters, HPC centers, and commercial cloud providers.

By joining their efforts EGI and OSG have opened their infrastructures to respond together to computational and data intensive COVID-19 research projects.

EGI and OSG COVID-19 call for projects

In alignment with many other organisations and project initiatives, EGI and OSG decided to join forces and provide specialised technical support, simulation tools and data intensive resources to promote international cooperation in research to tackle COVID-19. The call opened in April and is still receiving submissions.

EGI and OSG encourage COVID-19 research projects to apply to this call and take advantage of:

- sponsored access to computing and storage facilities of EGI and OSG to perform high-performance simulations and data analytics
- support from cloud, compute and data architects to port scientific applications and datasets to EGI compute and storage resources
- catalogue of ready-to-use, pre-configured virtual machine images that can run on any of the federated EGI cloud centres
- Jupyter notebook and other interfaces to access scientific data and to define, run and refine data analysis code and workflows

By applying to the call you can access:

- Infrastructure-as-a-Service clouds, HTC compute and storage resources of EGI and OSG in the USA
- Network of experts who can help you port custom data analytics and simulation applications and code to the infrastructure resources
- A portfolio of tools, applications and virtual machine images that can be used and customised on the infrastructure resources for your analysis
- Notebook interface, a workload manager and data management services to define and manage complex data staging and analysis scenarios
- Access to all these services via graphical portal, command line tools or Application Programming Interfaces

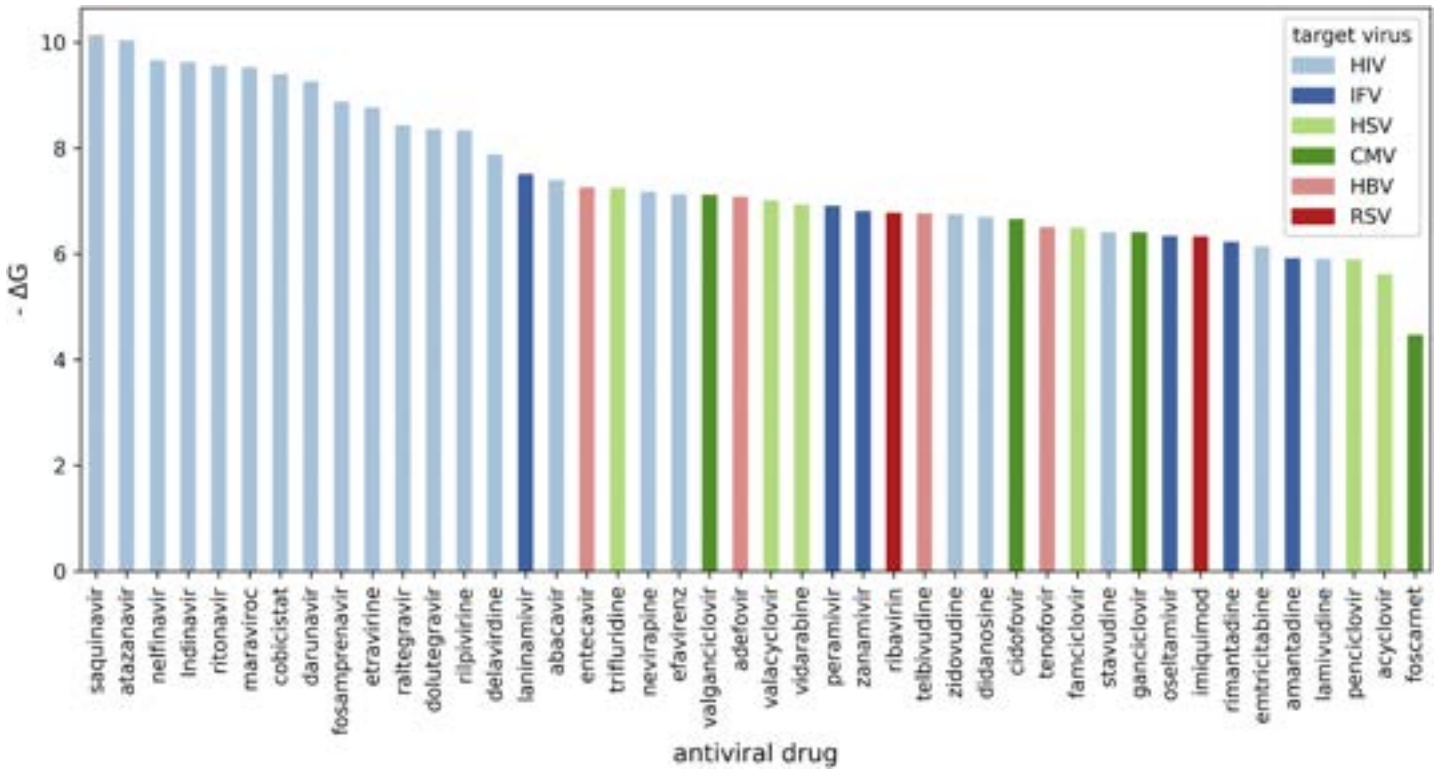
Use case

HIV COCKTAIL THERAPY TO COMBAT THE COVID-19 VIRUS

EGI services: Online Storage, Cloud Compute

Draško Tomić and his research team from the Rudjer Boskovic Institute, Croatia, have accessed EGI cloud resources to investigate possible new drug combinations. Spike glycoprotein is an essential element for the reproduction of the SARS-CoV-2 virus and its inhibition may pave new ways for treatment of COVID-19 patients. Due to this knowledge, Tomić and team analysed the inhibition of the glycoprotein with FDA-approved antiviral drugs and their double and triple combinations.

The team has used the VINI - a multi-drug multi-target docking service for COVID-19 - to perform a virtual drug screening, which currently runs on the supercomputer “Bura” at the Rijeka University. 44 approved generic antiviral drugs from the U.S. Food and Drug Administration (FDA) portal have been included in the screening process. As visualised in the figure below, the HIV drugs appear to have a binding energy to the SARS-CoV-2 glycoprotein higher than other antiviral drugs thus making them better candidates for investigating the impact of their double and triple combination on that virus.



Results of the effective inhibition of the SARS-CoV-2 spike glycoprotein has shown that it can slow or even completely influence the reproduction of the virus, increasing chances for the host immune system to fight it. The study provides scientific confirmation of the high efficacy of several combinations of approved HIV drugs. The research team notes, however, that a detailed pharmacokinetic analysis of the effects of individual drugs will be required to evaluate the benefits of such treatments or regimens, which goes beyond the scope of the study.

Thanks to the resources and the technical support received by EGI the team managed to finish their planned research in adopting the VINI model in finding best inhibitors for SARS-CoV-2 spike glycoprotein.

The full study and results are available in the paper "The screening and evaluation of potential clinically significant HIV drug combinations against the SARS-CoV-2 virus", published in Informatics in Medicine Unlocked journal, January 2021.

Use case

HOW PIG GENES ARE INVOLVED IN CORONAVIRUS INFECTIONS

EGI services: High-Throughput Compute, Online Storage

Many researchers have been looking into how the virus can spread from humans to animals. Scientists Luca Fontanesi and his colleagues at University of Bologna addressed the relationship between coronaviruses, human health, pig health and animal breeding in their most recent research. In practice, they investigated how pigs could represent a risk source of coronavirus infections for humans as this could cause damaging effects by disrupting livestock production chains.

The team took into account studies in humans that have shown how a host genome might confer different levels of resistance to coronavirus infections which can also be expected in animal species. In particular, their study focused on the genetic variability affecting four porcine genes that are directly involved in the progress of coronavirus infections. The analysis of variability was carried out for 25 porcine breeds, by mining different DNA databases and DNA sequencing datasets with a bioinformatics pipeline. Thus, the genetic variability was studied looking for genetic mutations affecting the protein encoded by these genes and by comparing this genetic information with that coming from humans.

The team's data mining identified over 2,200 DNA variants affecting four candidate genes of the pig genome that are known to be involved in coronavirus diseases. Up to 26% of the mutations were reported for the first time by their study. Furthermore, 29 variants resulted in altering the protein sequence. The comparative analysis with the human-related proteins resulted useful to identify this valuable animal model for defining genetic mechanisms associated with disease resistance. Many other genes have been also targeted in pigs and several other livestock species.

To support his research, Luca Fontanesi used computing resources offered by the EGI provider RECAS-BARI. The resources were used to replicate an improved bioinformatic pipeline for the identification of genetic variants, to download data and align DNA sequences, among others.

Luca's work can be considered a useful step towards conservation programs of pig genetic resources, concluding the following: 1) genetic resources could be reservoirs of host gene variability useful to design selection programs to increase resistance to coronaviruses and 2) the described variability in genes involved in coronavirus infections across many different pig populations might be part of a risk assessment including pig genetic resources.

The first results of the study are available in Luca's paper "Describing variability in pig genes involved in coronavirus infections for a One Health perspective in conservation of animal genetic resources", published in Scientific Reports Journal in February 2021.

Use case

HADDOCK HELPS SCIENTISTS TO ADVANCE COVID-19 RESEARCH

EGI services: High-Throughput Compute, Workload Manager, Online Storage, and Check-In

Talking about impact in the context of the novel COVID-19 research is still difficult in terms of finding a vaccine or any other solution to eliminate the virus. That is not to say that there isn't any progress being made, by plenty of researchers, many of them benefiting from open science initiatives.

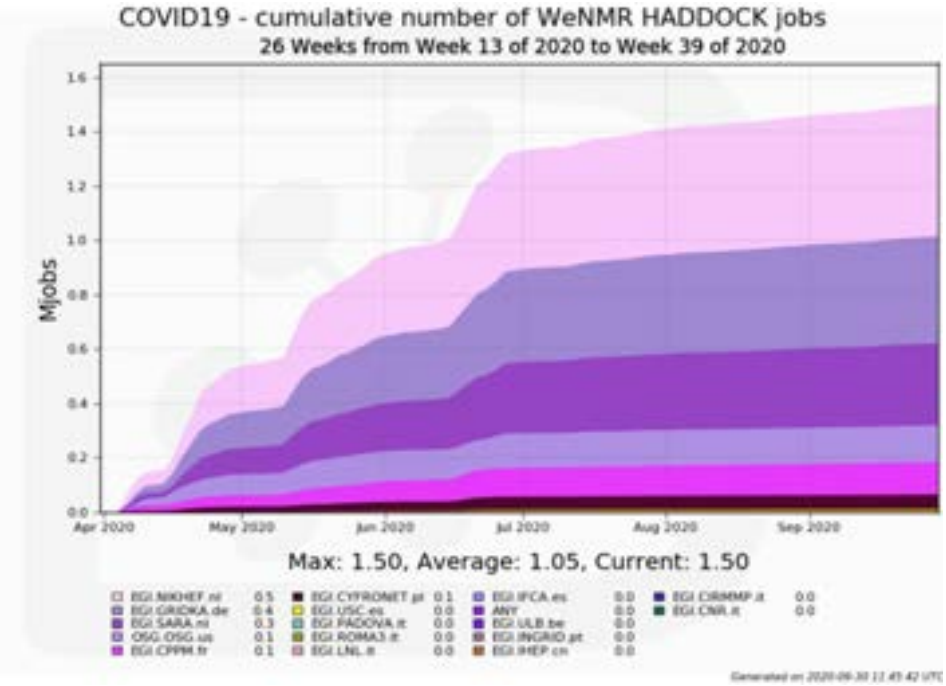
One example of these initiatives is HADDOCK – High Ambiguity Driven protein-protein DOCKing – one of the WeNMR software suites developed at Utrecht University and operated as a core software in the BioExcel Center of Excellence to help COVID-19 researchers to perform a rapid screening of thousands of chemical compounds against the protease structure.

From a technical perspective, HADDOCK supports complex simulations to generate 3D models revealing how virus proteins interact with human ones, or to dock small molecules to targets such as the SARS-CoV-2 protease.
In the context of the EGI and OSG COVID-19 initiative, the HADDOCK platform supported the drug repurposing effort by screening over 2000 chemical compounds against 3 COVID-19 related targets in up to 8 days using the computing resources jointly provided by EGI and OSG.

This potential drug target plays an essential role in processing the polyproteins that are translated from the viral RNA during replication. Targeting it and other viral proteins such as the RNA polymerase with small molecules could make for effective anti-coronaviral drug cocktails (as is done for example for HIV).

Due to the urgency, researchers made use of a strategy that repurposes existing and already approved drugs toward new diseases. The HADDOCK WeNMR platform supported this drug repurposing effort by screening over 2000 chemical compounds against three COVID-19 related targets in up to 8 days using EGI High-Throughput Compute resources.

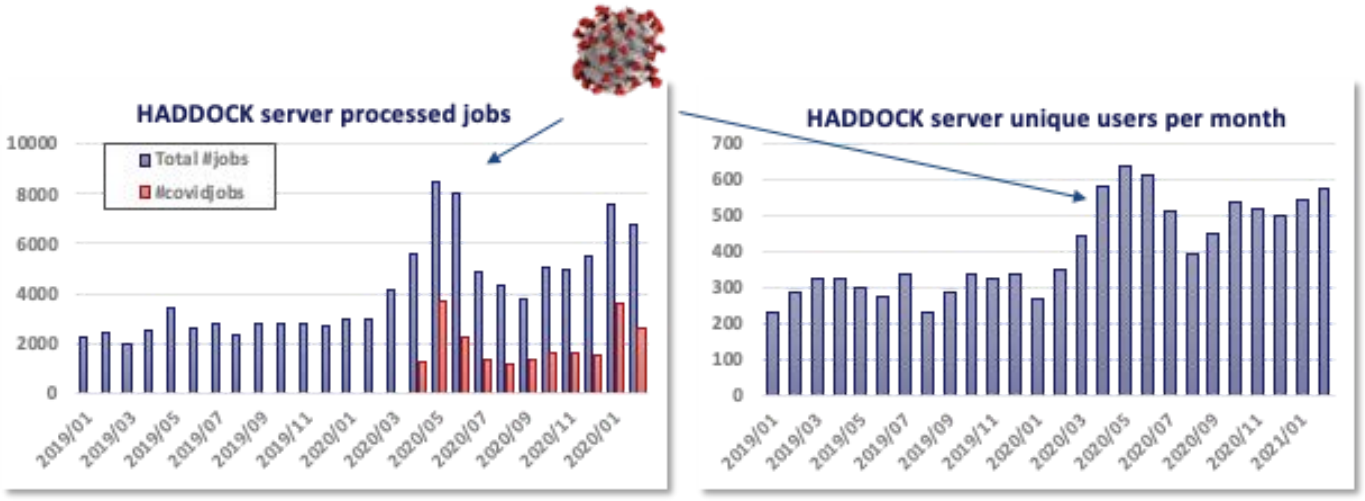
Thanks to the synergies and cooperation already established between EGI, the Open Science Grid and various high energy physics sites that committed to support COVID-19 related research, HADDOCK was able to more than double its processing capacity, serving on average 550 active users per months, who submitted over 11,000 simulations related to COVID-19 (the equivalent of ~1.5 million HTC jobs, ~2.7 million CPU hours) on the EGI and OSC grid resources over the months of April to September 2020.



Cumulative number of COVID-19 related HTC jobs on EGI resources for the period April-September 2020 (data source: DIRAC4EGI).

The initial results have revealed interesting compounds, some of which are already under clinical trials, supporting the validity of the screening methodology. A selection of these compounds is currently being experimentally tested by the Veterinary Faculty at Utrecht University.

In order to monitor COVID-19 related runs, the portal allows tagging of submission as COVID-19. Since this monitoring started, over 27,000 runs – not counting the drug repurposing screen – were processed to date. This tagging also allows to target the HTC jobs to sites specifically supporting this research. Another great result is to see the international mobilisation of additional resources to support HADDOCK jobs from both European, often high energy physics sites like Nikhef (Netherlands), CPPM (France), KIT (Germany) and USC-LCG2 (Spain), and the Open Science Grid.



HADDOCK server processed jobs and unique users per month. The red bars indicate COVID-related submissions.

Visit Bonvin Lab's website to check the progress of the research.

Contact information: Alexandre Bonvin - a.m.j.j.bonvin@uu.nl

Connect with us

www.egi.eu

/egi_einfra

/company/egi-foundation