

INFRAEDI-02-2018



BioExcel-2 Project Number 823830

## **D6.3 - High Performance Computing in support of COVID-19 Research**

*WP6: Management*



Copyright © 2019-2021 The partners of the BioExcel Consortium



This work is licensed under a  
[Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

## Document Information

<b>Deliverable Number</b>	D6.3
<b>Deliverable Name</b>	Deliverable Name
<b>Due Date</b>	2021-12-31 (PM36)
<b>Deliverable Lead</b>	KTH
<b>Authors</b>	Rossen Apostolov (KTH) Alexandre Bonvin (UU) Dmitry Morozov (JYU) Stian Soiland-Reyes (UNIMAN) Adam Hospital (IRB) Bert de Groot (MPG)
<b>Keywords</b>	Project Management
<b>WP</b>	WP6
<b>Nature</b>	Report
<b>Dissemination Level</b>	Public
<b>Final Version Date</b>	2021-12-14
<b>Reviewed by</b>	EB
<b>MGT Board Approval</b>	2021-12-21

## Document History

Partner	Date	Comments	Version
KTH	2021-11-16	First draft	0.1
UNIMAN	2021-11-21	Added WorkflowHub section	0.2
KTH	2021-12-14	Final edits	0.3
KTH	2021-12-17	Small corrections	0.4

## **Executive Summary**

The unexpected emergence and spread of the SARS-CoV-2 virus and COVID-19 disease had a dramatic impact on societies across the world. Very soon it became clear that concerted efforts are needed in the fight against the pandemic. As a focal point for the computational biomolecular research communities in Europe, BioExcel restructured its efforts in support of addressing the crisis. We developed and launched a number of portals for COVID-19 related data management and analysis. High-priority access to compute resources was provided by partner HPC centers. We started numerous collaborative research projects including close partnerships with pharma industry. Community support was extended. We participated in dedicated webinars and presentations to cover methodologies, experience and results from our work. We added a dedicated section to our website under category Research (<https://bioexcel.eu/covid-19-research/>) to present the various services and activities related to the disease. In this document we summarize the aforementioned activities.

## CONTENTS

<b>1</b>	<b>PORTALS AND ENVIRONMENTS .....</b>	<b>6</b>
	COVID-19 MOLECULAR STRUCTURE AND THERAPEUTICS HUB.....	6
	MAKING COVID-19 WORKFLOWS FAIR WITH WORKFLOWHUB.....	7
	BY-COVID CONSORTIUM.....	8
	BIOACTIVE COMPOUNDS AGAINST COVID-19 CHECKER .....	9
	BIOEXCEL-COVID-19 WEB-SERVER INTERFACE.....	9
	MEETING INCREASED DEMAND FOR HADDOCK.....	10
<b>2</b>	<b>PROJECTS .....</b>	<b>11</b>
	HIGH-THROUGHPUT FREE-ENERGY-BASED SCREENING OF ACE2 MUTATIONS.....	11
	WATER-TRIGGERED, IRREVERSIBLE CONFORMATIONAL CHANGE OF SARS-CoV-2 MAIN PROTEASE ON PASSING FROM THE SOLID STATE TO AQUEOUS SOLUTION .....	13
	LARGE-SCALE SARS-CoV2 MUTATION ANALYSIS USING BIOEXCEL HPC WORKFLOWS .....	13
	HPC EVALUATION OF SARS-CoV2 RdRp DRUG TARGETS USING MD SIMULATION.....	14
	STRUCTURE AND INTERACTION-BASED DESIGN OF ANTI-SARS-CoV-2 DNA APTAMERS .....	15
	HADDOCK DRUG REPURPOSING SCREEN USING EOSC HTC RESOURCES.....	16
	HADDOCK NMR-BASED MODELLING OF CRYPTIC SARS-CoV2 SPIKE WITH SUGAR INTERACTIONS .....	18
	EXSCALATE4COV CONSORTIUM .....	20
<b>3</b>	<b>EVENTS AND OUTREACH.....</b>	<b>21</b>
	WEBINARS .....	21
	CONFERENCES .....	21
	TRAINING COURSES .....	22
	HACKATHONS.....	22
	SOCIAL MEDIA AND WEBSITE.....	22
<b>4</b>	<b>PUBLICATIONS .....</b>	<b>23</b>
	REVIEW ARTICLE ON COVID-19 RESEARCH .....	23
	JOURNAL ARTICLES .....	23
	COMMUNITY LETTER: SHARING COVID-19 BIOMOLECULAR SIMULATION DATA .....	24
<b>5</b>	<b>FUTURE PERSPECTIVES.....</b>	<b>24</b>

## 1 Portals and environments

### COVID-19 Molecular Structure and Therapeutics Hub

*(Note: Presented also in D3.4 - User Community Support and Engagement Report (half-time update))*

We partnered with MolSSI - a US counterpart to our centre - to jointly create the BioExcel/MolSSI Molecular Structure and Therapeutics Hub. The Hub can be accessed at either <https://covid.bioexcel.eu/> or <https://covid.molssi.org/>. The BioExcel/MolSSI COVID-19 Hub is a community-driven data repository and curation service for molecular structures, models, therapeutics, and simulations related to computational research into therapeutic opportunities for COVID-19.

We worked with large and leading academic and commercial biomolecular research groups, including DE Shaw Research and some of the academic labs that are part of the Folding@Home consortium, to collect and share a wide range of data such as models of small molecules and large assemblies, simulation trajectory data, novel drug targets, etc., which may be useful to the wider community. The portal has quickly become a main reference data resource. It offers a variety of data storage and integration solutions, as follows:

- **GitHub** - the simplest storage solution for small molecular data
- **Zenodo** - larger data sets of up to 50 GB. We engaged directly with the Zenodo developers on facilitating access for the needs of the portal
- **Amazon Web Services** - special allocation has been granted for our COVID-19 data, with no specific limit on the size of the data sets
- **Fenix Infrastructure** (from the Human Brain Project, EU Flagship) - we started extensive discussions with the developers for data integration. We have preliminary discussions for adoption of the EBRAINS Knowledge Graph from HBP to facilitate data discovery on our Hub. This collaboration was supported by BioExcel researchers Adam Hospital, Modesto Orozco, Paolo Carloni, who are also part of HBP.
- **EMBL-EBI COVID-19 data portal** - we are in discussions to provide access to the BioModels and BioStudies resources

Thus, our portal provides flexibility of storage solutions while offering a single discovery entry point.

## Making COVID-19 workflows FAIR with WorkflowHub

The first public instance of the [WorkflowHub](https://covid19.workflowhub.eu/) registry was launched in April 2020, pushed ahead of schedule by the needs of the COVID-19 pandemic. An important aim of <https://covid19.workflowhub.eu/> is to help register and publicise known bioinformatics and cheminformatics workflows relevant to COVID-19, such as the virtual screening of the SARS-CoV-2 main protease.

The screenshot shows the WorkflowHub website interface. At the top, there's a navigation bar with 'WorkflowHub', 'Browse', 'Help', a search bar, and 'Register'/'Log in' links. Below this, the 'Workflows' section is active. A sidebar on the left contains filters for 'Query', 'Created At', 'Workflow Type', 'Tag', and 'Submitter'. The main content area shows 40 workflows matching the criteria 'Tag: covid-19'. The first workflow is 'Calculate the half maximal inhibitory concentration (IC50) for each compound used in a SARS-CoV-2 study' by OME. The second is 'COVID-19 sequence analysis on Illumina Amplicon PE data' by iwc. The third is 'COVID-19: variation analysis on ARTIC ONT data' by iwc. Each workflow entry includes a title, a brief description, the type (e.g., Jupyter, Galaxy), creator, submitter, and creation/last update dates.

**Figure 1:** <https://covid19.workflowhub.eu/> registers public COVID-19 workflows from a range of research groups across several major workflow management systems.

During the virtual [COVID-19 Biohackathon](#), in collaboration with ELIXIR, BioExcel-2, EOSC-Life and other volunteers, the development of the open source WorkflowHub prototype was significantly accelerated, together with a collective exercise to identify, curate and register public COVID-19 pipelines based on Galaxy, Nextflow, CWL, Snakemake, PyCOMPSs and other workflow languages.

Several research groups were contacted to assist in preparing their workflows for publishing, in addition to coordinating with other community efforts such as

<https://covid19.galaxyproject.org/> and curating existing public third-party workflows found in disparate repositories such as GitHub.

The WorkflowHub registry has by November 2021 [>40 public COVID-19](#) workflows (see figure 1) and has now matured to be in public beta, with >160 workflows registered overall across life science users from >70 research groups.

The WorkflowHub is listed in [FAIRsharing](#) and the [COVID-19 Data Portal](#) as a related resource. It has been a catalyst for community collaboration, bringing together workflow creators and workflow infrastructure developers in order to establish [FAIR Computational Workflows](#) as a principle and realized concept, expanding on just the download of a workflow definition to also record significant metadata, documentation, examples and to support continuous testing to [avoid workflow decay](#).

This move to enrich workflows to FAIR digital objects was driven initially by the COVID-19 needs and is now promoted more widely by BioExcel-2 partners and within EOSC. This work has also contributed to significant improvements to mature related community standards and best practices, including for [Bioschemas](#), [GA4GH Cloud](#) APIs, data packaging with [RO-Crate](#) and the [Common Workflow Language](#) (see BioExcel-2 deliverables [D3.5 – Best Practice Guides](#) and [D2.6 – Integration of BioExcel software framework with standards and international initiatives](#)).

Carole Goble; Stian Soiland-Reyes; Finn Bacall; Stuart Owen; Alan Williams; Ignacio Eguinoa; Bert Driesbeke; Simone Leo; Luca Pireddu; Laura Rodríguez-Navas; José M<sup>a</sup> Fernández; Salvador Capella-Gutierrez; Hervé Ménager; Björn Grüning; Beatriz Serrano-Solano; Philip Ewels; Frederik Coppens (2021): **Implementing FAIR Digital Objects in the EOSC-Life Workflow Collaboratory**. Zenodo <https://doi.org/10.5281/zenodo.4605654>

Niklas Blomberg & Katharina B. Lauer (2020): **Connecting data, tools and people across Europe: ELIXIR's response to the COVID-19 pandemic**. *European Journal of Human Genetics* 28 pp. 719–723. <https://doi.org/10.1038/s41431-020-0637-5>

## **BY-COVID consortium**

The WorkflowHub registry and the metadata community standard RO-Crate have a key role in the [BY-COVID project](#), launched in October 2021 following a Horizon Europe funding call HORIZON-INFRA-2021-EMERGENCY-01. This project has [53 European-wide partners](#) including EMBL-EBI, UNIMAN and BSC.

The core aim of the project is to ensure that data on SARS-CoV-2 and other infectious diseases can be found and used by everyone.



Planned work includes improving the FAIR metadata standards of the [COVID-19 Data Portal](#) (hosted by EMBL-EBI) as well as workflow execution through a *COVID-19 data platform* using WorkflowHub as registry (hosted by UNIMAN).

### **Bioactive compounds against Covid-19 checker**

The SARS-CoV-2 outbreak has prompted an unprecedented effort by the scientific community to find an effective vaccine or drugs to mitigate the viral infectiveness and symptoms, which is reflected in the over 4,000 publications that appeared in the last weeks. The volume of information is inaccessible by any single research group, and this may limit progress towards the rapid discovery of a COVID-19 therapy. Through a review of the most relevant scientific literature, and considering different levels of experimental evidence, we have identified over 150 compounds that are potentially active against COVID-19.

We have developed a [webportal](https://sbnb.irbbarcelona.org/covid19/) <https://sbnb.irbbarcelona.org/covid19/> to exploit this literature curation effort to identify other compounds with the potential to be effective against COVID-19. To this aim, we use the ChemicalChecker (CC), a resource that provides processed, harmonized and integrated bioactivity data for about 1M small molecules. In the current resource, we provide the results of a systematic similarity search across the large chemical space encompassed by the CC, thereby substantially expanding the portfolio of potential COVID-19 drug candidates.

### **BioExcel-COVID-19 web-server interface**

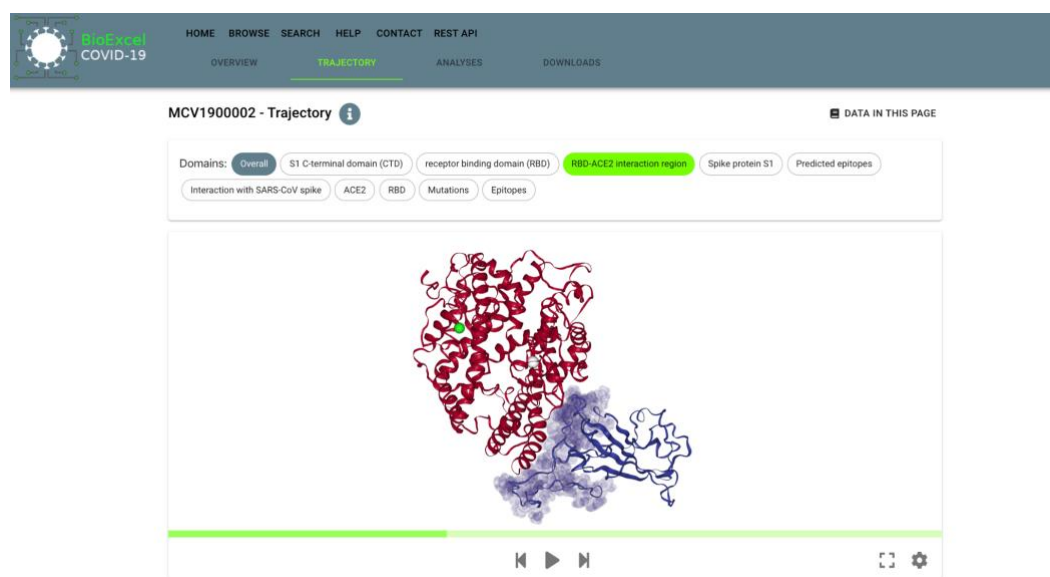
*(Note: Presented also in D3.4 - User Community Support and Engagement Report (half-time update))*

An online platform - <https://bioexcel-cv19.bsc.es> - has been launched to provide web-access to atomistic-MD trajectories for macromolecules involved in the COVID-19 disease. The project is part of the open access initiatives promoted by the world-wide scientific community to share information about COVID-19 research.

The BioExcel-CV19 database and associated web server (Fig. 2) includes interactive graphical representations of COVID-19 related MD trajectories and their corresponding analyses. The main objective of the BioExcel-CV19 project is to generate a tool for scientists interested in the COVID-19 research to interactively and graphically check key structural and flexibility features stemming from MDs.

As these features vary depending on the structure analyzed, specific analyses were performed, uploaded to the database, and represented in the web portal. These analyses and key features were collected by direct interaction with the authors of

the simulations. As an example, trajectories corresponding to the virus RBD-hACE2 complex include interface observables (e.g. residue distances, hydrogen bonds), allowing an easy analysis of their behaviour along the simulation.



**Figure 2:** BioExcel-CV19 web server screenshot. An RBD-hACE2 complex MD trajectory is shown, with the RBD-hACE2 interaction region highlighted.  
<https://bioexcel-cv19.bsc.es/#/browse/MCV1900002/trajectory>

All the analyses integrated in the web portal are completely interactive. Whenever possible, a direct link from the analysis to the 3D representation is offered, using NGL viewer tool. The fast extraction of a particular snapshot from the whole trajectory is possible thanks to the NoSQL Mongo database backend powering the server.

The whole set of trajectories (atomistic 3D coordinates for every atom and every frame of the simulation) and analyses are stored in this distributed database and efficiently retrieved on the fly from any web portal request. The large amount of data coming from the atomistic trajectories and their corresponding analyses is stored in the MongoDB distributed database.

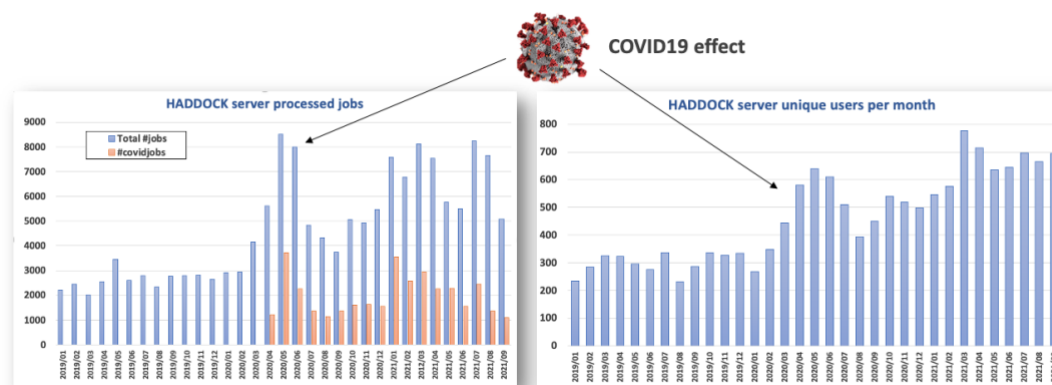
The entire pipeline is automated and new COVID-19 related trajectories are currently being processed. The set of analyses will be continuously extended, according to the suggestions of the authors of MD simulations.

### Meeting increased demand for HADDOCK

*(Note: Presented also in D3.4 - User Community Support and Engagement Report (half-time update))*

The [HADDOCK](#) portal, which enables the modelling of interactions between biomolecules, has seen since the beginning of the COVID-19 pandemic both an increase in user registration and use of the service. This is clearly seen in Figure 3

showing the number of submissions and unique users per month since the start of the project, with a clear increase from March 2020 onwards.



**Figure 3:** Number of processed jobs and unique users of the HADDOCK portal per month. The orange bars in the left plot correspond to COVID-19 related submission.

In order to meet the increased demand, changes have been made to the main HADDOCK Python code that manages the docking workflow, resulting in a decrease of the load on our servers, allowing us to more than double the number of concurrent jobs the server can handle.

We have also added to the portals a mechanism to tag submissions as being COVID-19 related; in order to both monitor such submissions, and also to redirect these jobs to HTC sites that have committed resources to COVID-19 projects (e.g. the US Open Science Grid, and several high energy physics sites in Europe like the Karlsruhe Institute of Technology and the Centre de Physique des Particules de Marseille). The increased usage of the HADDOCK portal is clearly visible in Figure 3 (right plot), showing the number of processed user submissions per month.

## 2 Projects

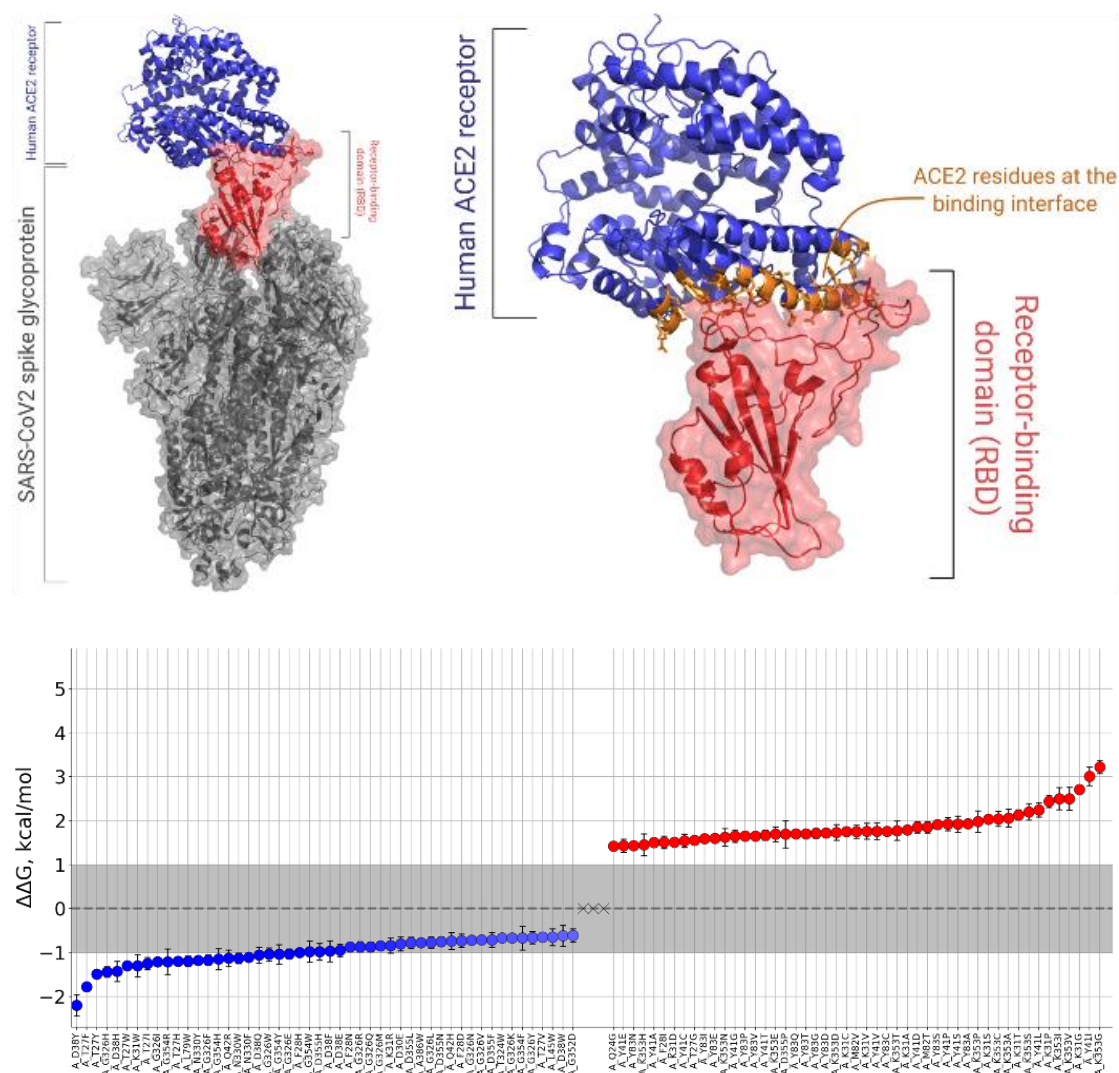
### High-throughput free-energy-based screening of ACE2 mutations

(Note: presented also in [D3.3 - Use Case Progress Report](#))

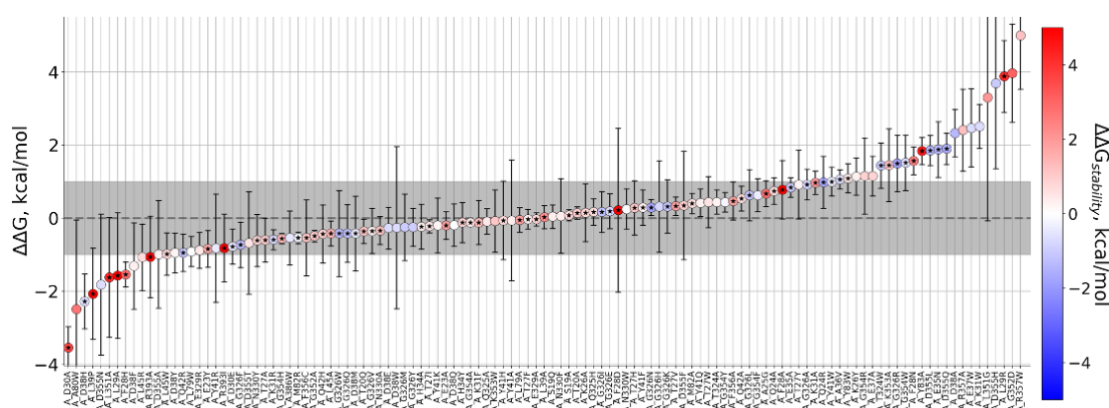
BioExcel partner MPG investigated the thermodynamics of the human ACE2 receptor interactions with the SARS-CoV2 spike protein receptor binding domain. For that purpose, a two tier strategy was employed for scanning the mutations of ACE2 in search of the candidates most stabilizing the complex as well as the apo receptor. In the first screening round, the Rosetta software was used for a high throughput screen of the complex interface residues. In total, more than 700 mutation induced binding free energy changes were estimated (figure 4). In the

second round, the most promising candidates were selected and evaluated with alchemical free energy calculations using PMX/GROMACS (figure 5).

All in all, a set of promising candidate mutations were identified that could subsequently be probed in an experimental setup by a (non-BioExcel) collaborator at Utrecht University. The findings will be presented as part of a concerted BioExcel manuscript that is in preparation for submission to WIREs Computational Molecular Science.



**Figure 4:** Site saturation mutagenesis by means of Rosetta. 700 mutations scanned.



**Figure 5:** Alchemical analysis of more than 100 mutations with GROMACS/PMX

### Water-Triggered, Irreversible Conformational Change of SARS-CoV-2 Main Protease on Passing from the Solid State to Aqueous Solution

This protein is a homodimer. Yet, a recent 0.1-ms-long molecular dynamics simulation performed by D. E. Shaw's research group [shows](#) that it readily undergoes a symmetry-breaking event on passing from the solid state to aqueous solution. As a result, the subunits present distinct conformations of the binding pocket.

By analyzing this long simulation, BioExcel partner PC in a collaboration with Prof. Parrinello's group uncovered a previously unrecognized role of water molecules in triggering the transition. Interestingly, each subunit presents a different collection of long-lived water molecules. Enhanced sampling simulations, along with machine learning approaches, further established that the transition to the asymmetric state is essentially irreversible.

Narjes Ansari, Valerio Rizzi, Paolo Carloni, Michele Parrinello (2021): **Water-Triggered, Irreversible Conformational Change of SARS-CoV-2 Main Protease on Passing from the Solid State to Aqueous Solution.** *Journal of the American Chemical Society* **143**(33) pp. 12930–12934. <https://doi.org/10.1021/jacs.1c05301>

### Large-scale SARS-CoV2 mutation analysis using BioExcel HPC workflows

(Note: presented also in [D3.3 - Use Case Progress Report](#))

Workflows designed and developed for Use Case 3 were quickly [modified](#) and updated by BioExcel partners IRB, NBD and BSC to tackle a particularly interesting quest: understanding how the virus has evolved by comparing its structure/genome to other coronavirus species and strains including GATg13, SARS-CoV as well as the US-variant. The gathered knowledge will be used to identify virus inhibition opportunities by means of *in silico* drug screening, starting from known and commercially available drugs first. The work is done in

collaboration with Prof. Roberto Burioni's laboratory for viral research at San Raffaele, Italy and Dr. Núria López-Bigas at the Institute for Research in Biomedicine, Spain.

The current project is focused on a) understanding the mechanism of virus entrance into the cell and the adaptation of the virus to different host species, b) understanding the different sensitivity to the virus (beyond the age) of the individuals, and c) predicting the next mutations of the virus and how it might adapt to be even more infectious.

The mechanism of entrance of the virus is based on one of the COVID-19 capsid proteins (Spike), which is recognized by an extracellular protease (ACE2). The main objective is to determine the impact of genetic changes in the viral *Spike Receptor Binding Domain* (RBD) and in the ACE2/RBD complex for the recognition of the virus.

We use free energy calculations to trace the impact of the mutational landscape on the binding of RBD to the host receptor proteins. BioExcel Use Case 3 workflows can help in decreasing the time needed to compute a large number of calculations efficiently using HPC resources. For the case of human ACE2, mutations identified as contributing most to binding free energy changes were investigated by MPG as described in the first study mentioned above.

An exascale project has recently started, aiming at computing all the combinations between the different known viral strains (RaTG13, SARS-CoV-2, SARS-CoV-1, SARS-CoV-2, and US-variant) against all the different known affected species (human, bat, zibeline, pangolin, etc.). The combinatorial explosion of the project clearly needs a massive HPC-focused workflow such as workflow 1 (WF1) developed for Use Case 3 (described in [D2.3 - First release of demonstration workflows](#)), together with a huge amount of computational power.

The project was presented to the PRACE COVID-19 Fast Track Call for Proposals, <https://prace-ri.eu/prace-support-to-mitigate-impact-of-covid-19-pandemic/>, and the study named "*Exploring Covid19 Infectious Mechanisms and Host Selection Process*", was awarded with 6 million core-hours on the Joliot-Curie Rome supercomputer (CEA/GENCI, France).

### **HPC evaluation of SARS-CoV2 RdRp drug targets using MD simulation**

(Note: presented also in [D3.3 - Use Case Progress Report](#))

BioExcel partner JYU modified the workflow implemented in Use Case 4 for analyzing fluorescent protein mutants with new parameters for evaluating the



available antiviral drugs that target the RNA dependent RNA polymerase (RdRp) of SARS-CoV2.

The objective of the project is to use MD simulations for RdRp in complex double stranded RNA (template-primer) and the nucleotide analogue inhibitors to provide the missing structural insights needed to: (a) understand the mode of action of the inhibitors, (b) predict the effect of chemical modifications of the nucleotide bases both inside the nucleotide binding site and into the RNA primer, (c) investigate possible cooperative effect of the inhibitor by modifying multiple bases in the primer strand, as these are speculated to induce premature primer chain termination with possible conformation changes in the RdRp complex.

In total, parameters and RdRp-inhibitor complex structure models were generated for 5 target nucleotide inhibitors (*Remdesivir*, *EIDD-2801*, *Favipiravir*, *Sofosbuvir*, and *Galidesivir*). Generated protein-RNA-inhibitor setups, force fields, and MD trajectories were uploaded to share publicly on the [joint MolSSI/BioExcel repository](#). The large-scale simulations of the RdRp complex were performed on HPC resources at CSC in Finland and at the Joliot-Curie Rome supercomputer (CEA/GENCI, France).

### **Structure and interaction-based design of anti-SARS-CoV-2 DNA aptamers**

*Aptamers* are artificial single-stranded RNA or DNA molecules consisting of 15–100 nucleotides. They can be designed to bind with high affinity and specificity to target proteins. The mechanism of their interaction with the target is similar to that of antibodies. It can be characterized as hydrophobic, electrostatic, hydrogen bonding, and van der Waals interactions, as well as base stacking and shape complementarity. Aptamers have a high potential in diagnostics, therapeutics, and drug delivery systems. However, they also could be used as antiviral drugs urgently needed during outbreaks of infections.

Modern aptamer selection process against novel infections is a complicated and time-consuming approach. Thus, efficient methods for rapidly screening sequences capable of blocking viruses binding to host cells should be developed beforehand. Synergy can be achieved by using computational methods together with experimental procedures to speed up an aptamer design. In this project we developed a reliable methodology for a rational *in silico et vitro* design of DNA aptamers to the SARS-CoV-2 Spike protein target. The new approach combines multiple steps in a workflow manner: (1) Molecular design, based on screening in a DNA aptamer library and directed mutagenesis to fit the protein tertiary structure; (2) 3D molecular modeling of the protein target; (3) Molecular docking of an aptamer with the target; (4) Molecular dynamics (MD) simulations of the complexes; (5) Accurate quantum-mechanical (QM) evaluation of the aptamer-

target interactions; (6) Experimental verification of structure and binding affinity using small-angle X-ray scattering, flow cytometry, and fluorescence polarization.

Using this new iterative design procedure, a highly specific aptamer to the receptor-binding domain (RBD) of the SARS-CoV-2 spike protein, was designed and validated with the experiments. Preliminary results demonstrate that the developed aptamers are promising candidates for detecting and/or blocking the SARS-CoV-2 virus.

This work has been done in a large collaboration consisting of groups from five countries all over the world. Expertise of the BioExcel in simulations of biological systems including proteins and DNA is a key component of that research. All simulations of the aptamer-RBD complexes were performed on HPC resources at CSC in Finland.

### **HADDOCK drug repurposing screen using EOSC HTC resources**

*(Note: initial results presented also in [D3.3 - Use Case Progress Report](#))*

Proteins that are crucial for the survival and replication of the virus are the most attractive targets for such studies. BioExcel partner UU focused on the SARS-CoV-2 main protease (3CLpro/Mpro) that plays an essential role in the virus replication process, the RNA dependent RNA polymerase (RdRp) and the human receptor ACE2.

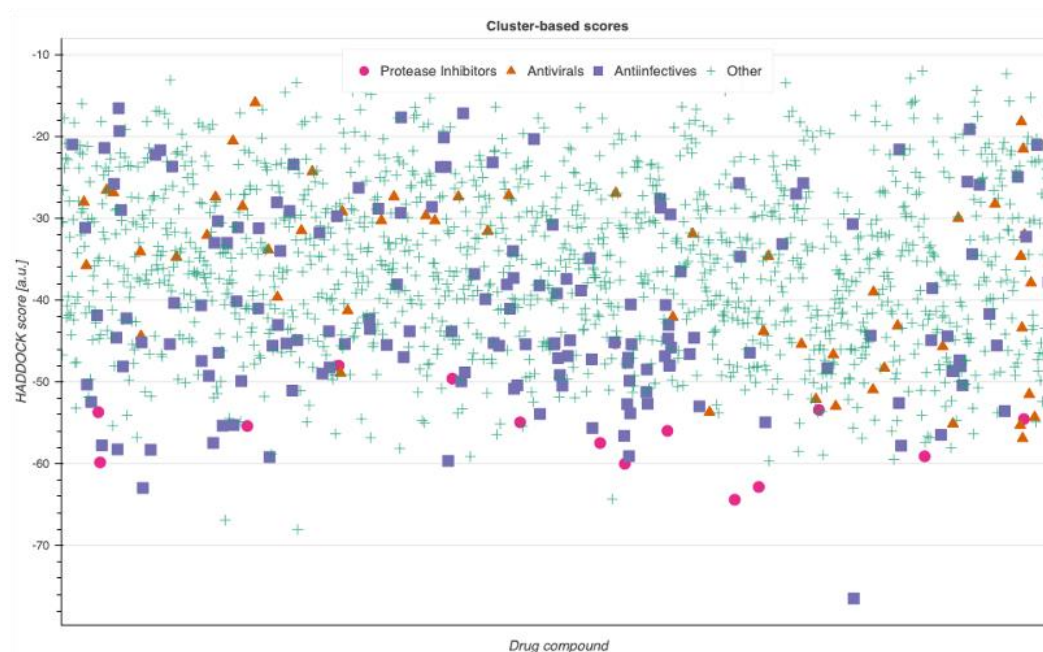
All three targets were screened against ~2000 approved drugs (and 6 experimental ones - these can be identified by searching the tables available at <https://www.bonvinlab.org/covid> using the term 'Investigational') against this particular protein. This was done using [HADDOCK2.4](#) following different strategies depending on the information available for each target. For 3CLpro a pharmacophore-based and a shape-based strategy adapted from our successful participation to the D3R Grand Challenges and described in [Protein-ligand pose and affinity prediction. Lessons from D3R Grand Challenge 3](#) [1]. This novel shape-based protocol was recently published [2].

For the other two targets a more "classical" HADDOCK approach was followed in which in which the residues lining the binding site were given to HADDOCK. More details of the protocol are provided on <https://www.bonvinlab.org/covid>.

The compounds were obtained from [Drugbank](#) and pre-processed using the [OpenEye](#) Omega software ([Hawkins et al. J. Chem. Inf. Model. 50 572-584 \(2010\)](#)). In addition to these compounds we also obtained some of their active metabolites from [PubChem](#). These can be recognised by the prefix CID. For the protease we used PDB entry [6Y2F](#) ([Zhang et al. Science, 2020](#)).



The screening of >2000 approved drugs ran in between 3 ½ and 8 days, depending on the target, using EOSC HTC resources. The results of our effort can be seen at <https://www.bonvinlab.org/covid> and below. More details of the protocol are provided on the same site and also highlighted below. For each target both cluster rankings and single structure rankings are provided. The scores (in arbitrary units) correspond to the HADDOCK score. An example is shown in Figures 6 and 7 below.



**Figure 6:** Overview of HADDOCK scores for the >2000 screened approved drugs against the main protease.

Show  entries

target	name	atc	score_pharm3D	score_tanimoto	best	rank	category	classification
DB09050	Ceftolozane	J01DI54	-76.472	-72.721	-76.472	1.0	Antiinfectives	Approved
DB00619	Imatinib	L01XE01	-53.026	-68.018	-68.018	2.0	Other	Approved
DB00460	Verteporfin	S01LA01	-66.905	-54.423	-66.905	3.0	Other	Approved
DB05521	Telaprevir	J05AP02	-56.025	-64.398	-64.398	4.0	Protease-Inhibitors	Approved
DB01259	Lapatinib	L01XE07	-64.314	-57.024	-64.314	5.0	Other	Approved
DB13947	Testosterone-enantate-benzilic-acid-hydrazone	NA	-63.336	-56.418	-63.336	6.0	NA	Approved
DB00303	Ertapenem	J01DH03	-47.774	-62.967	-62.967	7.0	Antiinfectives	Approved
DB06290	Simeprevir	G01AE10 J05AP05	-62.839	-45.632	-62.839	8.0	Protease-Inhibitors	Approved
DB11979	Elagolix	NA	-61.135	-51.301	-61.135	9.0	NA	Approved
DB01319	Fosamprenavir	J05AE07	-60.014	-45.474	-60.014	10.0	Protease-Inhibitors	Approved

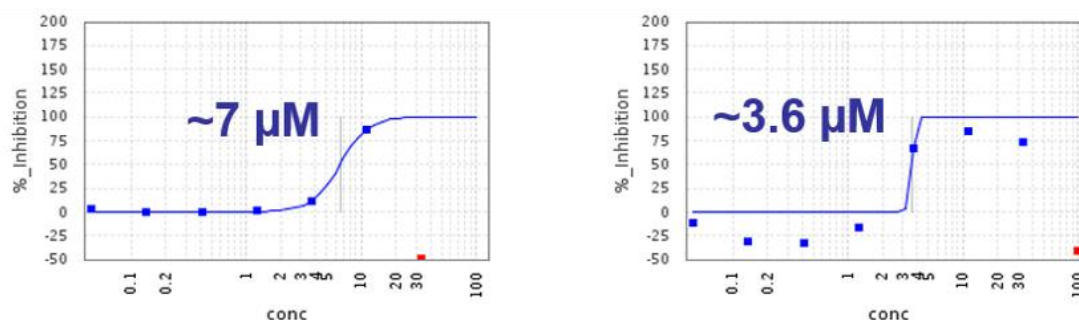
Showing 1 to 10 of 2,020 entries

Previous  2 3 4 5 ... 202 Next

**Figure 7:** Overview of the top 10 compounds from the HADDOCK screen of approved drugs. The table is searchable and available online at: <https://www.bonvinlab.org/covid>

The plots show one data point per compound. In total there are 2020 compounds, grouped into one of four categories 'Protease Inhibitors', 'Antivirals', 'Anti-infectives' and 'Other'. The first is made up of compounds that are known to inhibit proteases, the second antiviral medications, the third general anti-infectives and the last everything else. The 'NA' group corresponds to compounds that have no specific associations - these include dietary supplements, amino acid residues, etc, which are not shown in these plots but are listed in the [online tables](#).

We reached out to collaborators in the [CARE-IMI](#) consortium to have our most promising compounds experimentally tested. The experiments yielded several compounds with activity against the protease in the micromolar range but subsequent testing revealed them to also be cytotoxic, rendering them less relevant for clinical applications. Two of the most promising compounds were *imatinib* and *lapatinib*, listed as #2 and #5, respectively in our cluster-based ranking (Figure 8).



**Figure 8:** Inhibition assays of Imatinib (left) and Lapatinib (right).

The infection assays, affinity measurements and toxicity assays were carried out in the groups of Frank Kuppeveld at Utrecht University and Johan Neyts at KU Leuven.

A presentation about this work was given at the BioExcel 2020 virtual summer school and is [available online](#) on the BioExcel YouTube channel.

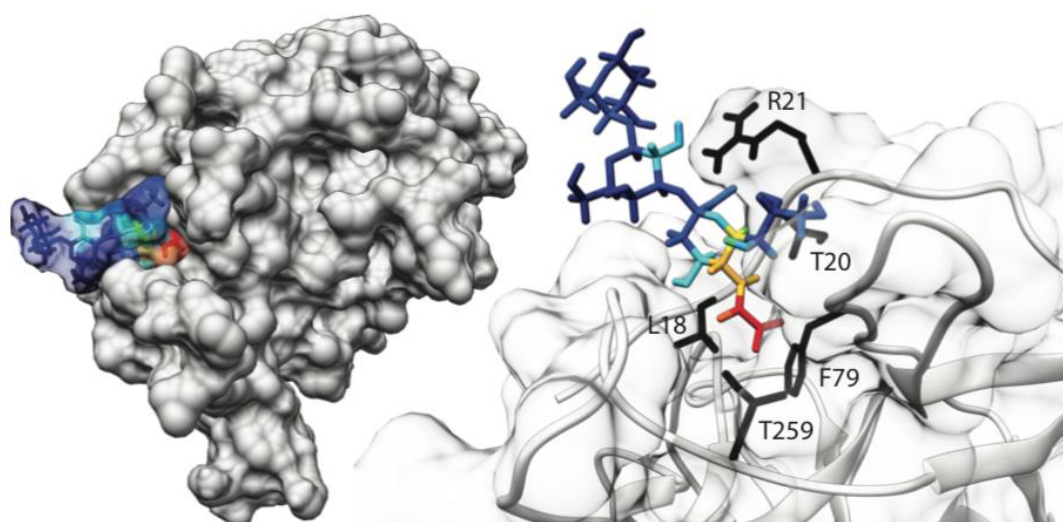
### **HADDOCK NMR-based modelling of cryptic SARS-CoV2 spike with sugar interactions**

The surface proteins found on both pathogens and host cells mediate entry (and exit) and influence disease progression and transmission. Both types can bear host-generated post-translational modifications such as glycosylation that are essential for function but can confound biophysical methods used for dissecting key interactions. Several human viruses (including non-SARS-coronaviruses)

attach to host cell-surface N-linked glycans that include forms of sialic acid (sialosides).

There remains, however, conflicting evidence as to if or how SARS-associated coronaviruses might use such a mechanism. In *B-origin-lineage-SARS-CoV-2* spike trimer 'end on'-binding to sialoside sugars was revealed contrasting with 'extended surface'-binding for heparin sugar ligands.

In collaboration with groups in Oxford we generated using HADDOCK with NMR saturation transfer data (uSTA) a structural model of the Sars-Cov2 spike N-terminal domain (NTD) with various glycans (Figure 9).



**Figure 9:** HADDOCK NMR-STD-based model to 2,3- $\alpha$ -sialyllactose binding to the NTD domain of Sars-Cov2 spike protein.

Consistent with this, uSTA-glycan binding was minimally- perturbed by antibodies that neutralize the ACE2-binding domain (RBD) but strongly disrupted in spike from the *B1.1.7/alpha* and *B1.351/beta* variants-of-concern, which possess hotspot mutations in the NTD. Sialoside binding in *B-origin-lineage-NTD* was unequivocally pinpointed by cryo-EM to a site that is created from residues that are notably deleted in variants (e.g. H69,V70,Y145 in alpha).

An analysis of beneficial genetic variances in cohorts of patients from early 2020 suggests a model in which this site in the NTD of *B-origin-lineage-SARS-CoV-2* (but not in alpha/beta-variants) may have exploited a specific sialylated-poly lactosamine motif found on tetra-antennary human N-linked-glycoproteins in deeper lung.

Together these confirm a novel binding mode mediated by the unusual NTD of SARS-CoV-2 and suggest how it may drive virulence and/or zoonosis via modulation of glycan attachment. Since cell-surface glycans are widely relevant to biology and pathology, uSTA can now provide ready, quantitative, widespread

analysis of complex, host-derived and post-translationally modified proteins with putative ligands relevant to disease even in previously confounding complex systems.

This work is available as a preprint [3].

---

[1] P.I. Koukos, L.C. Xue and A.M.J.J. Bonvin. [Protein-ligand pose and affinity prediction. Lessons from D3R Grand Challenge 3](#). *J. Comp. Aid. Mol. Des.* 33, 83-91 (2019).

[2] P.I. Koukos, M.F. Reau and A.M.J.J. Bonvin. [Shape-restrained modelling of protein-small molecule complexes with HADDOCK](#). *J. Chem. Inf. and Mod.* 61, 4807–4818 (2021).

[3] C.J. Buchanan, B. Gaunt, P.J. Harrison, A. Le Bas, A. Khan, A.M. Giltrap, P.N. Ward, M. Dumoux, S. Daga, N. Picchiotti, M. Baldassarri, E. Benetti, C. Fallerini, F. Fava, A. Giliberti, P.I. Koukos, A. Lakshminarayanan, X. Xue, G. Papadakis, L.P. Deimel, V. Casablanco-Antràs, T.D.W. Claridge, A.M.J.J. Bonvin, Q.J. Sattentau, S. Furini, M. Gori, J. Huo, R.J. Owens, A. Renieri, [GEN-COVID Multicenter Study](#). J.H. Naismith, A. Baldwin, B.G. Davis. [Cryptic SARS-CoV2-spike-with-sugar interactions revealed by 'universal' saturation transfer analysis](#). *bioRxiv* <https://doi.org/10.1101/2021.04.14.439284> (2021).

### Exscalate4Cov Consortium

*(Note: Presented also in D3.4 - User Community Support and Engagement Report (half-time update))*

KTH and BSC are partners in the newly established [Exscalate4Cov](#) consortium bringing GROMACS and HPC consultancy expertise to tailor molecular dynamics simulations for large-scale executions. NBD is screening its proprietary virtual library ChemistriX as well as other open libraries (Zinc) and contributes altruistically to the project.

JYU started a number of projects on PRACE and CSC clusters, which are directly connected to COVID-19, and works closely with the research groups of Petri Pihko (expert in organic synthesis) and Varpu Marjomäki (virology specialist).

We have also used BioExcel's position to facilitate connections between the E4C consortium and research groups from EU13 member states, in particular establishing collaborative contacts with academia and the pharmaceutical industry in Bulgaria.

JULICH, as partner of the consortium, started several projects on inhibitors of the viral main protease with the groups A. Zaliani (in vitro essays, Fraunhofer) and Paola Storici (X-ray crystallography, Elettra Synchrotron, Trieste).

### 3 Events and Outreach

(See also [D4.3 – Progress report and update on training and dissemination plan](#))

#### Webinars

- [“HPC And Bigdata Approaches In Covid-19 Research”](#), Modesto Orozco  
*CECAM Webinars: CoVid-19: challenges and responses in simulation, modeling and beyond*
- [“The Proteins of SARS-CoV2-2, joint data repositories, and community collaborations”](#), Erik Lindahl  
*CECAM Webinars: CoVid-19: challenges and responses in simulation, modeling and beyond*
- [“Using Icei Resources For Atomistic Molecular Dynamics Simulations - A Covid-19 Example”](#), Modesto Orozco  
*Fenix Infrastructure Webinar*
- [“Bioexcel Building Blocks And HPC. A Test Case In Covid Research”](#), Modesto Orozco  
*CECAM Webinars: The importance of being H.P.C. Earnest*
- [“Drug repurposing against SARS-Cov2 using HADDOCK”](#), Alexandre Bonvin  
*Interdisciplinary consortia for the study of pandemics, CIC biomaGUNE*
- [1st EU-ASEAN webinar on HPC-COVID19 related modeling and diagnostic](#), Rossen Apostolov
- [“RO-Crate: Describing and packaging FAIR Research Objects”](#), Stian Soiland-Reyes, Carole Goble (2021), [Scottish Covid-19 Response Consortium](#)

#### Conferences

- “Data portals for Covid19 simulation efforts”, Modesto Orozco  
*CECAM Conferences on Covid19 research. Invited Lecture. Lausanne, 2020*
- “Big Data and HPC to fight COVID-19”, Modesto Orozco  
[VIII Jornada de Bioinformàtica i Genòmica](#), Institut d’Estudis Catalans, Barcelona, 2020
- [“Towards an efficient use of exa-scale high-performance computing \(COVID-19 as a case example\)”](#), Modesto Orozco  
*Keynote Lecture, IEEE eScience 2021, Innsbruck, 2021*
- [“Towards an Efficient Use of Exa-Scale High-Performance Computing”](#), Modesto Orozco  
*Invited Speaker, SC2021*
- “BioExcel-CV19: a database of COVID-19 related Molecular Dynamics trajectories”, Adam Hospital  
*CECAM Workshop: [Open Databases Integration for Materials Design](#), 2020*



- “[BioExcel-CV19: a database of COVID-19 related Molecular Dynamics trajectories](#)”, Adam Hospital  
CECAM Mixed-gen Session 4: Data Driven Science, 2020
- “FEP-Guided Discovery of Potent Inhibitors of the Main Protease of SARS-CoV-2”, William Jorgensen, EMBO Workshop: Advances and Challenges in Biomolecular Simulations, 18 - 21 October 2021

### Training courses

The following COVID-19 short talks were delivered during the BioExcel Seasonal schools and are available through the BioExcel YouTube pages:

- [BioExcel Summer School 2020 on Biomolecular Simulations](#), short talk on [HADDOCK-based drug repurposing screening for COVID-19](#), Alexandre Bonvin, UU.
- [BioExcel Winter School 2020 on Biomolecular Simulations](#), short talk ['Using GROMACS and pmx with the BioExcel Building Blocks \(BioBB\) library to tackle COVID-19 research'](#), by Adam Hospital, IRB

### Hackathons

- [COVID-19 Biohackathon](#) including work on RO-Crate, WorkflowHub and Bioschemas

### Social media and website

- Dissemination of COVID-19 research webpage <https://bioexcel.eu/covid-19-research/>
- <https://instruct-eric.eu/haddock-screen-of-2000-approved-drugs-for-covid19>
- <https://eosc-portal.eu/news/haddock-support-covid-19-research>
- COVID-19 focused [webinars](#) on social media and newsletters since April
  - <https://bioexcel.eu/webinar-summarschool-2021/>
  - <https://bioexcel.eu/webinar-covid-19-molecular-structure-and-therapeutics-hub-2020-06-25/>
  - <https://bioexcel.eu/webinar-computationally-designing-therapeutic-antibodies-combining-immune-repertoire-data-and-structural-information-2021-11-09/>
- <https://bioexcel.eu/bioexcel-center-of-excellence-in-support-of-covid-19-research/>
  - The tweet advertising this blog on BioExcel’s work on COVID-19 Research achieved over 14,000 impressions, making it our most popular tweet so far.

## 4 Publications

### Review article on COVID-19 research

In addition to this deliverable, we have prepared a review article on COVID-19 research, which will be submitted to WIREs Computational Molecular Science (<https://wires.onlinelibrary.wiley.com/journal/17590884>)

### Journal Articles

- P.I. Koukos, M.F. Reau and A.M.J.J. Bonvin. [Shape-restrained modelling of protein-small molecule complexes with HADDOCK](https://doi.org/10.1021/acs.jcim.1c00796). *J. Chem. Inf. and Mod.* 61, 4807–4818 (2021). <https://doi.org/10.1021/acs.jcim.1c00796>
- C.J. Buchanan, B. Gaunt, P.J. Harrison, A. Le Bas, A. Khan, A.M. Giltrap, P.N. Ward, M. Dumoux, S. Daga, N. Picchiotti, M. Baldassarri, E. Benetti, C. Fallerini, F. Fava, A. Giliberti, P.I. Koukos, A. Lakshminarayanan, X. Xue, G. Papadakis, L.P. Deimel, V. Casablanca-Antràs, T.D.W. Claridge, A.M.J.J. Bonvin, Q.J. Sattentau, S. Furini, M. Gori, J. Huo, R.J. Owens, A. Renieri, GEN-COVID Multicenter Study, J.H. Naismith, A. Baldwin, B.G. Davis. **Cryptic SARS-CoV2-spike-with-sugar interactions revealed by ‘universal’ saturation transfer analysis**. *bioRxiv* <https://doi.org/10.1101/2021.04.14.439284> (2021).
- V. Mironov I.A. Shchugoreva, P.V. Artyushenko, D. Morozov, N. Borbone, G. Oliviero, T.N. Zamay, R.V. Moryachkov, O.S. Kolovskaya, K.A. Lukyanenko, Y. Song, I.A. Merkuleva, V.N. Zabluda, G. Peters, L.S. Koroleva, D.V. Veprintsev, Y.E. Glazyrin, E.A. Volosnikova, S.V. Belenkaya, T.I. Esina, A.A. Isaeva, V.S. Nesmeyanova, D.V. Shanshin, A.N. Berlina, N.S. Komova, V.A. Svetlichnyi, V.N. Silnikov, D.N. Shcherbakov, G.S. Zamay, S.S. Zamay, T. Smolyarova, E.P. Tikhonova, K.H.-C. Chen, U. Jeng, G. Condorelli, V. de Franciscis, G. Groenhof, C. Yang, A.A. Moskovsky, D.G. Fedorov, F.N. Tomilin, W. Tan, Y. Alexeev, M.V. Berezovski and A.S. Kichkailo. **Structure and Interaction Based Design of Anti-SARS-CoV-2 Aptamers**. (2021) *under review*
- Juan Aranda, Modesto Orozco. **RNA-Dependent RNA Polymerase From SARS-CoV-2. Mechanism Of Reaction And Inhibition By Remdesivir**. *bioRxiv* 2020.06.21.163592; doi: <https://doi.org/10.1101/2020.06.21.163592>
- Tiziana Ginex, Clara Marco-Marín, Miłosz Wieczór, Carlos P. Mata, James Krieger, Maria Luisa López-Redondo, Clara Francés-Gómez, Paula Ruiz-Rodríguez, Roberto Melero, Carlos Óscar Sánchez-Sorzano, Marta Martínez, Nadine Gougéard, Alicia Forcada-Nadal, Sara Zamora-Caballero, Roberto Gozalbo-Rovira, Carla Sanz-Frasquet, Jeronimo Bravo, Vicente Rubio, Alberto Marina, The IBV-Covid19-Pipeline, Ron Geller, Iñaki Comas, Carmen Gil, Mireia Coscolla, Modesto Orozco, José Luis Llácer, José-Maria Carazo. **The structural role of SARS-CoV-2 genetic**

**background in the emergence and success of spike mutations: the case of the spike A222V mutation.** bioRxiv 2021.12.05.471263; doi:

<https://doi.org/10.1101/2021.12.05.471263>

- Adam Hospital, Modesto Orozco. **Evolutionary Path and Host-selection Mechanism of SARS-CoV-2.** (*in preparation*)

### **Community Letter: Sharing COVID-19 Biomolecular Simulation Data**

*(Note: Presented also in D3.4 - User Community Support and Engagement Report (half-time update))*

Sharing of data is vital to the urgent need of lead identification for therapies, diagnostics, and vaccines for COVID-19. To maximize the impact, BioExcel members have signed the international [community letter](#) which aims to connect scientists and improve communication between simulation, experimental and clinical data investigators.

Rommie E. Amaro, Adrian J. Mulholland (2020): **Community Letter Regarding Sharing Biomolecular Simulation Data for COVID-19.** *J. Chem. Inf. Model.* **60**(6) pp. 2653–2656. <https://doi.org/10.1021/acs.jcim.0c00319>

## **5 Future perspectives**

Two years have passed since the pandemic outbreak. Despite the rapid development of effective vaccines and extensive work on medicine against the disease, the world is still fighting to get control over ever emerging new mutations of the virus. BioExcel will continue operating existing, and developing new, infrastructures such as portals, servers, environments, dedicated HPC resources. We will continue to lead and participate in research projects on the topic, as well as to increase efforts towards education of computational scientists in the usage of HPC.