# EGI-ACE Open Call no.1

Checkpoint meeting with Shepherds

## AMBER-based modelling of SARS-CoV-2 Spike protein

Shepherd: Doina Cristina Duma/INFN

**Dissemination level:**

**Disclosing Party:**

**Recipient Party:**

EGI-ACE receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 101017567.

# Outline - Max 10’ long talk

- *Background about the scientific use case*
- *Ambition, Impact and Challenges*
- *Integration Support*
- *Capacity Requirements*
- *Timeline*

# Background about the scientific use case

## *AMBER-based modelling of SARS-CoV-2 Spike protein (1)*

Overview of the scientific community representing the Data Space provider/Early Adopter use case.

- **Title:** ***"High-throughput atomistic glycan shield model of Fully-glycosylated Full-length SARS- CoV-2 Spike Protein in a Viral Membrane provides insights to design spike protein inactivators."***

Team members/organisations involved, location and type(s) of users

- **Principal Investigator**: Pasqualina D'Ursi, **Institute of Biomedical Technologies - National Research Council of Italy**
- Aditional contacts: Andrea Manconi, Alessandro Orro
- Representing an heterogeneous group of more than 11 researchers (bioinformatics, biotechnology, engineering, physics and computer science.) and 2 PhD students
- **Organizations**:
  - Institute of Biomedical Technologies of the National Research Council (**CNR-ITB**)
    - **Bioinformatics Laboratory**
  - University of Brescia (**UniBs**)
    - **Macromolecular Interaction Analysis Unit** (MIAU) in the Department of Molecular and Translational Medicine
    - New collaboration aimed at the design of compounds endowed with the capacity to block/inactivate the spike protein of the SARS- CoV2 virus
    - Doctorate project - "Design and validation of virucidal compounds for the development of anti-SARS-CoV2 Personal Protective Equipment"
  - University of Genoa (**UniGe**)
    - **Department of Pharmacy**
    - Doctorate project - "In silico methods applied on druggable proteins to identify transient pockets: new approaches for studying drug-target molecular mechanisms. A case study on CFTR."

# Background about the scientific use case

*AMBER-based modelling of SARS-CoV-2 Spike protein (2)*

- **CNR-ITB,** expertise in
  - The field of **molecular modelling and drug discovery**
  - The Implementation and maintenance of specialized **bioinformatics infrastructures**

- **MIAU - UniBs**, expertise in
  - study of ligand/receptor and drug/target interactions of physiological, pathological and pharmacological relevance
  - mutual exchange of **experimental data and computational predictions** allowed the deep comprehension of the mechanism of action(s) of various biological molecules and **drugs** in the fields of **cystic fibrosis and viral infections**

- Dept. of Pharmacy – **UniGe,** expertise in
  - The field of **drug discovery and drug design**
  - **Application** of different **computational strategies** for the rational design of ligands targeted to Cystic Fibrosis Transmembrane conductance Regulator (**CFTR**), Trace Amine Receptors (TAARs 1), serotoninergic receptors (5HT1) and PDE4 inhibitors

# Ambition, Impact, Challenge(s)

*Ambition*

- *Scientific:*
  - **development of compounds** able to inactivate the spike protein by evaluating the occupancy of the spike protein by glycans and identification of glycan holes, thus providing opportunities for inactivators to bind
  - **Perform a network analysis** on the glycan shield of spike protein to find the glycans that are most important for an effective shield.
  - Use the information collected in this study for the **repositioning of known drugs** in order to identify putative antiviral drugs and to design anionic polymers both targeted to the spike glycan holes
- Computing
  - **perform 17 molecular dynamics simulations** for a total of 8.5 microseconds of compound-spike protein complexes selected in a previous analysis performed by the research teams
  - Simulations starting from detailed dynamic of the **fully**-glycosylated **full-length** SARS-CoV-2 Spike protein, obtained from a well known repository [1]
  - **Use clustering, Solvent-accessible surface area (SASA) and network analysis** in order to identify glycan holes and obtain the protein conformations for drug repositioning and anionic polymers design

*Impact*

- Results obtained through the implementation of the proposed use case will be the **basis for the implementation of specific research studies aimed at identifying antiviral drugs and anionic polymers** targeted to inactivate the Spike protein
- use the knowledge acquired in this project to **implement a specific research activity** aimed at better characterize and investigate the glycanproteins complexity
- **training** in **drug discovery** studies and in **advanced biosimulations** of the two PhD students
- **Share** the molecular dynamic trajectories resulting from simulations with the scientific community **through public repositories**.

*Challenges*

- Although GPU devices have successfully been used to address different researches, **they have proved to be inadequate to efficiently deal with the use case in hand**. Main limitations are related to their **computation power** and the **memory size** that do not permit to fit the entire computation on the GPU memory, reducing notably the performance.
- accessing more modern GPUs will help to evaluate their computational power in order to **update the existing infrastructure** in the near future.

# Integration Support

- *Containerization of the analysis workflow*
- *Access to provided resources – EGI Check-in,*
  - *New research community or use an already available VO with specific sub-group*
  - *Possibility to enroll/register non IDEM users (PhD students)*
- How to use of the EGI Data transfer service, in particular for the output transfers
- ***No expertise on cloud computing***
  - access training material and/or specific training events

# Capacity Requirements

Requested - access to **modern and powerful GPU** devices as well as **suitable storage and transfer data services**

**10 virtual machines (VMs)**
- 7 VMs dedicated to run dynamic simulations on known drug protein complexes
- 3 VMs to simulate dynamics of polyanionic polymers protein complexes

The VMs should be configured as follow:
- Image: **Centos 7** image configured to run Docker containers leveraging NVIDIA GPUs (available in the EGI applications database);
- Flavour: **2 GPU Nvidia T4, 8 cores (for GPU), 24GB RAM**;
- GPU device**: Nvidia T4 16GB** (provided by CESNET);
- Block storage: it is required a different volume for VM dedicated to simulate dynamics of (A) known drug protein complexes with respect those dedicated to simulate dynamics of ((B) polyanionic polymers protein complexes:
    - (A): **volume of 5TB for each VM**
    - (B): **volume of 130TB for each VM**
- **EGI Data transfer service:** volumes should be exposed to the EGI transfer service to allow easy sharing of large sets of files resulting from simulations (order of TB)

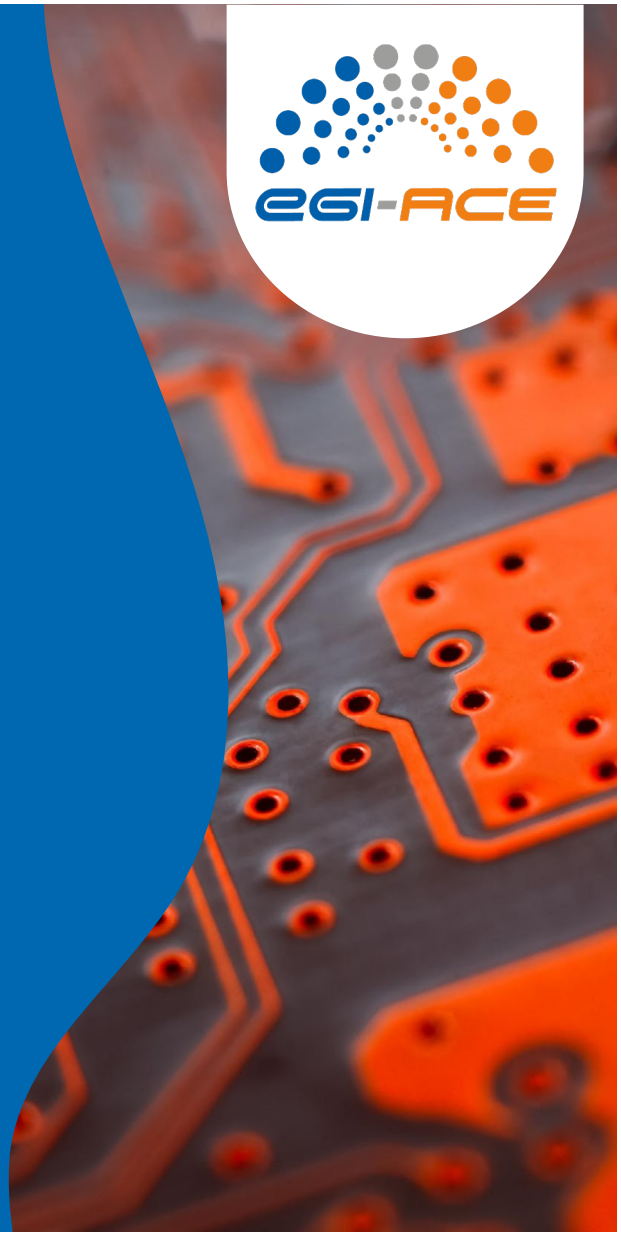
# Timeline

*“Considering a public benchmark on AMBER20 with NVIDIA T4 GPUs we estimated in 67 days the time required for a simulation of (A), and 105 days for simulations of (B)”*

Initial considerations for the timeline**(*)**:

- **Day 1**: Environment configuration
- **Day 4**: Starting molecular protocol setup and dynamic simulations on the first 7 of known drug protein complexes and on 3 polyanionic polymers protein complexes.
- **Day 71**: Completing simulations of first 7 known drug protein complexes.
- **Day 72**: Starting molecular protocol setup and dynamic simulations on the second set of 7 known drug protein complexes
- **Day 109**: Completing simulations on polyanionic polymers protein complexes
- **Day 139**: Completing simulations on second set of known drug protein complexes.

(*) - timeline does not take into account the time to download resulting output files (460TB)


EGI-ACE

# Thank you!

Contact: **egi-ace-po@mailman.egi.eu**
Website: **www.egi.eu/projects/egi-ace**

**EGI Foundation**

**@EGI_eInfra**

EGI-ACE receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 101017567.