Towards a three dimensional mapping of virus-host proximal protein interactions

Dr Yorgos Sofianatos, Dr Étienne Coyaud, Dr Caroline Demeret

What are the mechanisms of action of the SARS-CoV-2 virus inside an infected cell?
Combining modern screening methods sensitive to the proximity of protein interactions with computational tools for network layouts, a first picture of the viral invasion pattern is starting to emerge.

Adobe Stock © sdecoret

The COVID-19 pandemic has mobilised the scientific community worldwide in an unprecedented race to understand all aspects of the novel SARS-CoV-2 virus, ranging from its fundamental biology to strategies against infection.

This large-scale effort received support from governments and funding agencies in many countries, with resources being directed towards studies and actions that could help reduce the huge burden of the pandemic on human societies.

In early 2020, an international collaboration combining the expertise of several European and North American teams was formed to shed light on the interactions of the invading virus within human cells. The project was spearheaded by the French National Institute of Health and Medical Research (INSERM) PRISM Laboratory, Lille; the Biomedical Sciences Research Center (BSRC) "Alexander Fleming", Athens; and Pasteur Institute, Paris. Financial support was received from the French National Agency for Research (ANR) COVID-19 Flash funding scheme. Two Marie Skłodowska-Curie Individual Fellows at the PRISM and "Alexander Fleming" laboratories shifted priorities from their main projects to join forces and pursue this urgent goal.

The project's approach is based on a combination of modern techniques from interactomics (the branch of biology studying the complex web of protein interactions in living organisms) and algorithms for the visualisation of graphs (mathematical objects modelling interconnected systems). Together, they enable the creation of a comprehensive 3D chart of the virus-host proteins interplay taking place inside the host cell's volume.

While maps (*interactomes*) of protein-protein interactions (PPIs) have been around for a long time, they consist of only abstract representations of protein organisation in the cell. The PPIs identification method we implemented (BioID) delineates the population of host proteins (*interactors*) in the immediate proximity of each viral protein (bait) in living cells. Hence, it opens up the possibility of estimating actual relative bait-prey distances, creating the *proximal interactome*.

The premise is that the identification of proteins within a 10–20 nm spherical radius around a bait protein can be depicted through a new kind of geometric map, revealing regional enrichment. Indeed, by positioning baits and interactors relative to each other in a data-defined reference frame, we define occupation volumes in close proximity to viral baits in living cells populated by proteins likely to be involved in the infection process. This conceptually novel approach aims at pointing to unknown mechanisms of virus invasion and possible strategies to counteract them.

A proximity-dependent biotinylation identification (BioID) technique was performed at PRISM Laboratory. It relies on an enzymatic reaction driven by a catalytic moiety fused to each viral bait and proteins in living cells. An advantage is that this reaction covalently labels neighbouring host proteins with biotin. This implies that the preservation of PPIs upon the purification process is no longer required since biotinylated species can be captured on streptavidin columns and identified by mass spectrometry. We thus employ harsh lysis conditions to solubilise all cell proteins and reconstitute all proximal interactomes, probing a kind of 'host cell neighbourhood' of each of the virus' 27 bait proteins. The extracted proximal protein information provides the basis for reconstructing their geometric layout inside the cell space.

In addition to BioID, the complementary nanoluciferase two-hybrid (N2H) method was used to assay the direct binding of selected proximal interactions. With a methodology developed and performed at Pasteur Institute, about 30 per cent of the tested protein pairs were orthogonally confirmed and annotated as direct.

Overall, this dual approach of combining global BioID with direct N2H protein screening ensured a high level of validation and resulted in a comprehensive dataset of more than 10 000 high confidence virus-host protein interactions. Once the protein interaction data have been collected and quality controlled, it is the task of computer algorithms to translate it into a meaningful 3D representation. In our approach, utilising computational tools developed at BSRC "Alexander Fleming", the data is used to construct a large graph (or network, as it is commonly called)—essentially a set of pairwise links of varying strengths, capturing all the proximity relations detected through BioID approaches.

In aggregating all the links, the challenge is to reconstruct the host proteins' locations and determine their relative distances and volumes as faithfully as possible. For this purpose, computational techniques for the geometric layouts of graphs are useful. These algorithms can ingest a graph and, by way of simulating a physical analogue model until it reaches equilibrium, arrive at a set of final 3D coordinates for each of the graph's nodes (representing the viral and host proteins in our case). This type of algorithm, called force-directed layouts, showed good results regarding expected protein neighbourhoods known from other studies, providing validation for this approach.

In other words, probing the host cell with viral 'bait' proteins and detecting local neighbourhoods brings us full circle—via the computational modelling detour described previously—to an approximation of those protein locations in the cellular space.

To make the 3D map usable and informative, it is crucial to expose the data in a form that enables exploration and analysis for extracting useful hypotheses. For this purpose, we have built a comprehensive dashboard, available at **www.sars-cov-2-interactome.org**.
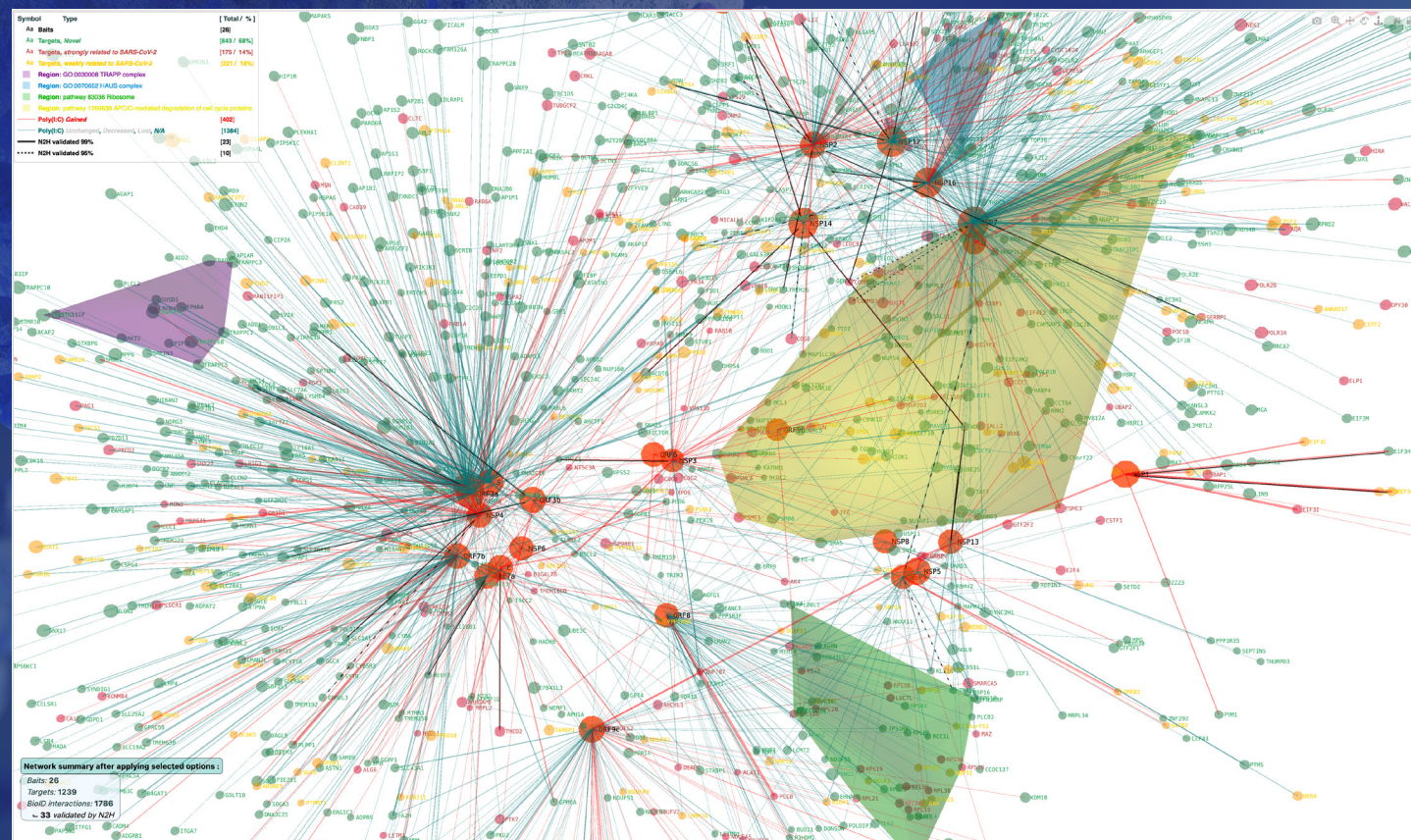
Figure 1: A view of the interactome taken from the dashboard webpage (www.sars-cov-2-interactome.org), highlighting the locations related to APC/C-mediated degradation of cell proteins (yellow region), ribosome (green region), TRAPP complex (purple region) and HAUS complex (blue region). Baits are denoted by larger orange nodes and host interactors by smaller nodes of different colours depending on novelty. Red links denote interactions appearing under conditions mimicking viral infection, while black links denote interactions validated by N2H.

The full interactome network dataset is provided, annotated with protein metadata (function, location, biological role) and augmented with results from previous studies regarding SARS-CoV-2 reported interactions.

An interactive 3D visualisation of the interactome is the main ingredient of the dashboard. Users can navigate within their browser, zooming and rotating in space, giving an intuitive illustration of the relative localisations of proteins inside the cell volume.

Through the dashboard functionality, one can focus on subnetworks of the interactome using multiple criteria. These include filtering by degree (a protein's total number of interactions) and novelty of interactor or type of interaction. Additionally, it is possible to focus on groups of proteins participating in specific biological processes, pathways

or cellular components and highlight their respective 3D localisation regions. This unprecedented level of customisation is important in reducing complexity and isolating manageable parts of the interactome. The interactive visualisation reflects the user selections at any given time, offering an intuitive view of the particular subset of data under consideration (Figure 1).

Through the study of the interactome, we have managed to identify interesting clues regarding the SARS-CoV-2 activity upon invasion. Amongst others, they include:

- the targeting of multiple proteins participating in the host's inflammatory response and cytokine production, which could affect the balance between immune response efficacy and progression to hyperinflammation—an important factor for the overall outcome of the disease

- interference with the cell's odorant and taste receptors, indicating links to anosmia and taste agnosia, both common COVID-19 symptoms

- interplay with the interferon signalling pathway, a possible mechanism of escape of the virus from the host's immune response

- hints to gender-dependent suppression of the host's mechanisms for viral detection, potentially causing a higher risk of severe symptoms in men, another known aspect of COVID-19 disease.

Overall, our interactions dataset contains a wealth of information that can provide new insights about how the SARS-CoV-2 virus hijacks and remodels the host protein system. The combination of complementary BioID and N2H techniques provides a valuable pipeline for identifying relevant protein

interactions in the cells. We believe that the availability of the interactome dashboard to the scientific community at large will encourage further exploitation of this rich dataset, as we plan to integrate additional layers of information from other published sources in the future.

The 3D modelling methodology is novel and promising for recreating the cell's spatial configuration while also having some limitations in its current form. Modelling the proteins as points in space is a simplification of the real picture: these are molecules moving inside the cell and are found in varying numbers of copies, modified forms and in multiple locations, often simultaneously. Therefore the point locations can only be indicative of the centroid of an extended region where a protein might be found. A more sophisticated modelling by a diffuse protein location (or probability cloud density) is something we are looking into. Another future direction is focusing our computational approach on comparisons of the proximal interactomes between normal and infected cells, which could uncover patterns of spatial occupancy reshaping due to viral activity.

As more proximal proteomics data are published, a data-driven approach of 3D positioning of proteins within the cell volume is opening up a new field towards differential mapping of proteins between normal and infected cells. We envision this novel approach as a new therapeutic path to target, not just specific proteins but cell regions appearing as central in structuring pathological networks.