EuroHPC Summit Week 2022
#PRACEdays

HPC workflow compartmentalises a multiscale simulation aimed at the personalisation of COVID-19 treatments and facilitates its scalability

Vincent Noël[1,*], Javier Conejero[2,*], Jose Carbonell[2], Miguel Vázquez[2], Gaurav Saxena[2], Miguel Ponce de Leon[2], Thaleia Ntiniakou[2], David Vicente[2], Emmanuel Barillot[1], Laurence Calzone[1], Rosa M Badia[2], Alfonso Valencia[2,3,$] and Arnau Montagud[2,$]

1 Institut Curie, INSERM U900, Mines ParisTech, University PSL, France
2 Barcelona Supercomputing Center, Spain
3 ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain
(*) co-first authors
($) co-last authors

As a result of the unprecedented COVID-19 pandemic, researchers now have a vast amount of new publications that cover different clinical and molecular aspects of the disease. In spite of this, many underlying molecular events that lead SARS-CoV-2 infection to the observed clinical traits of COVID-19 remain unclear.

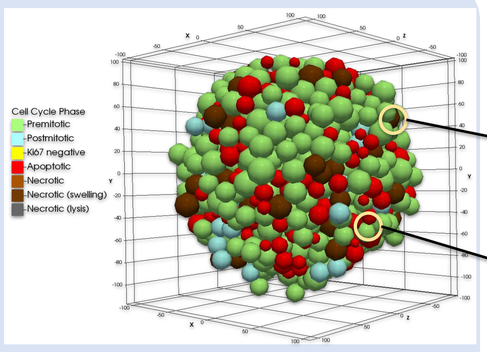
We hereby present the use of a multiscale modelling framework, termed PhysiBoSS[1], that integrates MaBoSS[2], a stochastic Boolean modelling software, into PhysiCell-COVID[3,4] to leverage of cell- and pathway-specific Boolean models for the study of patient-specific COVID infections. However, these frameworks usually lack the potential to consider many use cases simultaneously in workflows.

We have used PyCOMPSs[5] to orchestrate the patient-specific mechanistic exploration of COVID-19 infection of thousands of patients in scalable workflows that use Singularity containers[6] as building blocks.

PhysiCell

Multiscale modelling frameworks prove useful in integrating mechanisms that have very different time and space scales, as in the study of viral infection, human host cell demise and immune cells response.

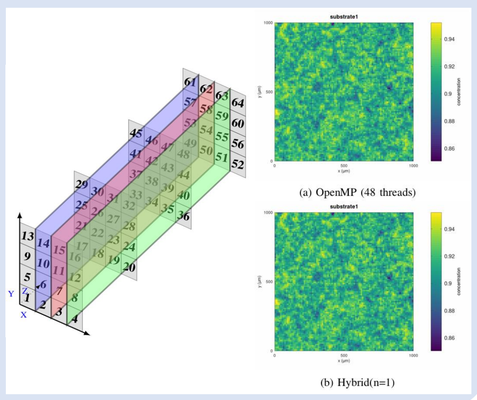
PhysiCell[3] is a multiscale multicellular agent-based tool with the highest potential to be expanded and built upon. It is used for the modelling of the population of cells and can help understand the role of the interactions between cells and their environment.



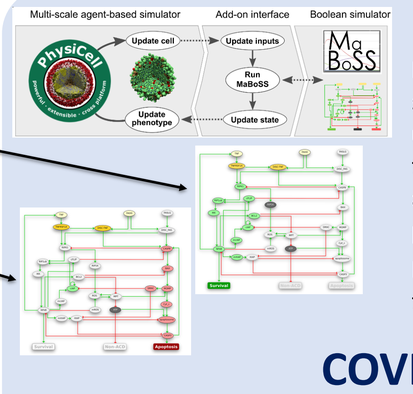
PhysiCell-X

PhysiCell[3] supports only shared-memory parallelization using OpenMP. We have restructured PhysiCell to support distributed parallelism using MPI and parallelize the generic core kernels of PhysiCell including simulation initialization, domain partitioning and agents' generation[7].

Preliminary results of PhysiCell-X expand the scope of the simulations by several orders of magnitude and enable the simulation of complex behaviours with different types of cells, substrates, drugs or three-dimensional domains.



PhysiBoSS



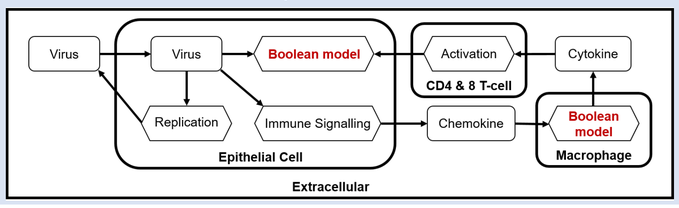
PhysiBoSS[1] merges PhysiCell[3] and MaBoSS[2] and allows for the combined study of genetic and environmental perturbations in many different set-ups. The use of **MaBoSS**, a tool that uses Monte-Carlo kinetic algorithm to perform **continuous time stochastic simulations on logical models**, allows for semi-quantitative evaluation of the model's phenotypes and perturbations.

PhysiBoSS was used on top of the PC4COVID initiative[4] that use PhysiCell to study the dissemination of the SARS-CoV-2 virus on lung epithelia.

COVID use case in PhysiBoSS

The use of PhysiBoSS on COVID offers mechanistic insights of SARS-CoV-2 infection and dissemination among human host cells. To obtain these models, we have taken advantage of the COVID-19 Disease Maps (C19DM)[8] which integrates and formalises mechanistic knowledge using current systems biology standards.

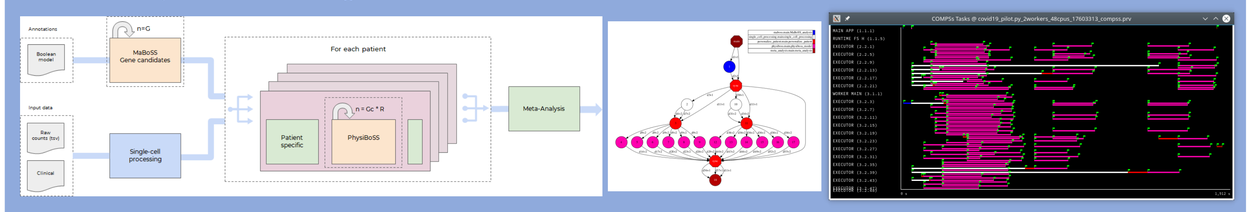
For instance, we worked with an apoptosis model for epithelial cells and a macrophages model from C19DM, modified it to connect it to the agent-base model and identified different mutants that affect apoptosis and immune cells' response.



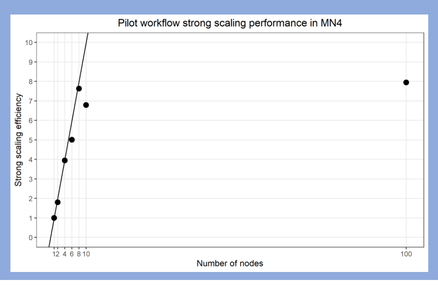
The framework integrates virus infection, epithelial host cell demise and different immune cells' response.

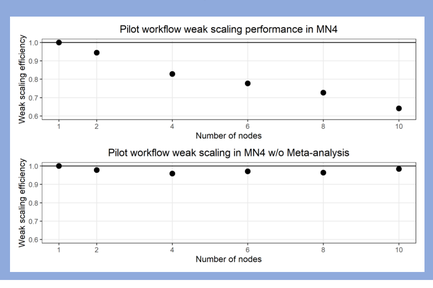
Workflow management for HPC

We devised a workflow that compartmentalises the tasks in building blocks, such as analysing single-cell data using the *seurat* R package, using single-cell data to tailor the Boolean models to their cell types, identifying interesting interventions, running the simulations and evaluating the simulations' results. This workflow was run for each patient for up to a total of 2000 patients and in replicates to take care of the intrinsic stochasticity of the modelling tools.

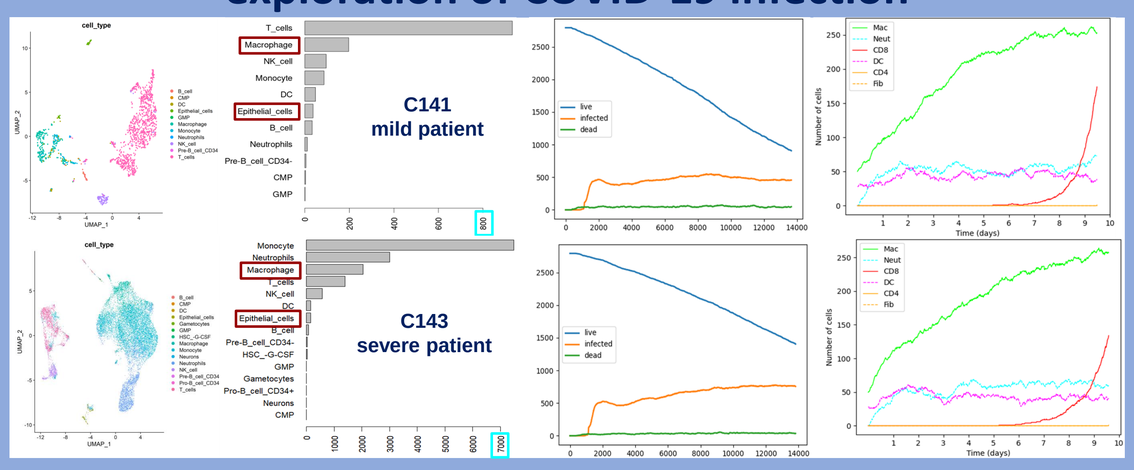


We tested the scalability of this workflow in MareNostrum4 with notable results. The strong scaling is dependent on the scaling of the different building blocks considered. The weak scaling works very well for the parallel building blocks, and its losses are constrained to the few linear parts of the workflow.





PhysiBoSS allows for the patient-specific mechanistic exploration of COVID-19 infection



With PhysiBoSS we can inspect COVID19 patient-specific perturbations that affect epithelial cells' apoptosis, immune cells' response and heterogeneous cell populations

Perspectives

- Fully integrate MPI to scale up the simulations using more HPC cores
- Refactor the linear parts of the workflow to increase the overall performance
- Include models for other immune cells

References

1. Letort *et al.*, PhysiBoSS: a multi-scale agent-based modelling framework integrating physical dimension and cell signalling. Bioinformatics. 2019, bty766
2. Stoll *et al.*, MaBoSS 2.0: an environment for stochastic Boolean modeling. Bioinformatics. 2017, 33, 2226–2228.
3. Ghaffarizadeh *et al.*, PhysiCell: An open source physics-based cell simulator for 3-D multicellular systems. PLOS Computational Biology. 2018, e1005991
4. Getz *et al.*, Rapid community-driven development of a SARS-CoV-2 tissue simulator. bioRxiv. 2020
5. Tejedor *et al.*, PyCOMPSs: Parallel computational workflows in Python, The International Journal of High Performance Computing Applications, 31(1), 66–82. 2017
6. Kurtzer *et al*. Singularity: Scientific containers for mobility of compute. PLoS ONE. 2017, 12(5), e0177459
7. Saxena *et al.*, BioFVM-X: An MPI+OpenMP 3-D Simulator for Biological Systems. Computational Methods in Systems Biology, Lecture Notes in Computer Science, 2021, 266–279.
8. Ostaszewski *et al.*, COVID19 Disease Map, a computational knowledge repository of virus–host interaction mechanisms. Molecular Systems Biology. 2021, 17, e10387.

Contact: arnau.montagud@bsc.es, @ArnauMontagud

Funding

This work is part of the European HPC/Exascale Centre of Excellence in Personalized Medicine (PerMedCoE, http://permedcoe.eu/, H2020-ICT-951773), whose purpose is to adapt cell-level simulation tools to supercomputing environments and to provide an easy-to-use interface to systems biology end users.

The EuroHPC Summit Week conference series receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823767 (PRACE-6IP) and 824151 (HPC-GIG).

#EHPCSW