

Lessons learned from urgent computing in Europe: Tackling the COVID-19 pandemic

Núria López^{a,b}, Luigi Del Debbio^{a,c}, Marc Baaden^{a,d}, Matej Praprotnik^{a,e,f,1}, Laura Grigori^{a,g}, Catarina Simões^a, Serge Bogaerts^a, Florian Berberich^{a,h}, Thomas Lippert^{a,h}, Janne Ignatius^{a,i}, Philippe Lavocat^{a,j}, Oriol Pineda^{a,k}, Maria Grazia Giuffreda^{a,l}, Sergi Girona^{a,k}, Dieter Kranzlmüller^{a,m}, Michael M. Resch^{a,n}, Gabriella Scipione^{a,o}, and Thomas Schulthess^{a,l}

Edited by Dan Stanzione, The University of Texas at Austin, Austin, TX, and accepted by Editorial Board Member William H. Press September 10, 2021 (received for review December 22, 2020)

PRACE (Partnership for Advanced Computing in Europe), an international not-for-profit association that brings together the five largest European supercomputing centers and involves 26 European countries, has allocated more than half a billion core hours to computer simulations to fight the COVID-19 pandemic. Alongside experiments, these simulations are a pillar of research to assess the risks of different scenarios and investigate mitigation strategies. While the world deals with the subsequent waves of the pandemic, we present a reflection on the use of urgent supercomputing for global societal challenges and crisis management.

urgent computing | COVID-19 pandemic | high-performance computing | Partnership for Advanced Computing in Europe | PRACE COVID-19 fast track

In this perspective we present an analysis of various scientific efforts undertaken in Europe to combat the COVID-19 pandemic based on advanced and computationally intensive numerical simulations that require the use of large-scale computers. The five largest European supercomputing centers supported some of these efforts with free-of-charge computational resources, through the Partnership for Advanced Computing in Europe (PRACE).*

This paper presents a number of computational models and methods relevant to the fight against the COVID-19 pandemic, including biostructural studies, docking and screening for pharmaceutical applications, computational fluid dynamics to study droplet spread and transmission of the virus, organ modeling

to support diagnosis and treatment of the disease, and epidemiology. We discuss the use of artificial intelligence (AI) in these areas and the readiness of the codes to run on high-performance computers.

The COVID-19 pandemic began between late 2019 and early 2020 in the city of Wuhan, Hubei, China. Identification of the virus occurred relatively quickly. The virus's genetic profile and the crystallographic blueprint of its spike proteins were published as open science to the community within weeks, providing the crucial structural information needed to develop biomodels[†] that ultimately led to messenger RNA vaccine formulation (1). At the same time, scientific analysis of the disease began rapidly at different levels:

^aPartnership for Advanced Computing in Europe 1050 Bruxelles, Belgium; ^bInstitute of Chemical Research of Catalonia, The Barcelona Institute of Science and Technology 43007 Tarragona, Spain; ^cSchool of Physics and Astronomy, The University of Edinburgh, Edinburgh EH9 3FD, United Kingdom; ^dLaboratoire de Biochimie Théorique, CNRS, Unité Propre de Recherche 9080, Université de Paris, 75005 Paris, France; ^eLaboratory for Molecular Modeling, National Institute of Chemistry, SI-1001 Ljubljana, Slovenia; ^fDepartment of Physics, Faculty of Mathematics and Physics, University of Ljubljana, SI-1000 Ljubljana, Slovenia; ^gSorbonne Université, CNRS, Université de Paris, Inria, Laboratoire Jacques Louis Lions, Alpines, 75012 Paris, France; ^hInstitute for Advanced Simulation, Jülich Supercomputing Centre D-52425 Jülich, Germany; ⁱCSC-IT Center for Science, 02101 Espoo, Finland; ^jGrand Équipement National de Calcul Intensif F-75015 Paris, France; ^kBarcelona Supercomputing Center 08034 Barcelona, Spain; ^lSwiss National Supercomputing Centre 6900 Lugano, Switzerland; ^mLeibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities 85748 Garching, Germany; ⁿHigh-Performance Computing Center Stuttgart 70569 Stuttgart, Germany; and ^oSupercomputing Applications and Innovation Department, Cineca Consorzio Interuniversitario, 40033 Casalecchio di Reno, Italy

Author contributions: C.S. collected and provided data; S.B., F.B., T.L., J.L., P.L., O.P., M.G.G., S.G., D.K., M.M.R., G.S., and T.S. provided supercomputer time; N.L., L.D.D., M.B., M.P., and L.G. analyzed data; and N.L., L.D.D., M.B., M.P., L.G., C.S., S.B., F.B., T.L., J.L., P.L., O.P., M.G.G., S.G., D.K., M.M.R., G.S., and T.S. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. D.S. is a guest editor invited by the Editorial Board.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](#).

¹To whom correspondence may be addressed. Email: praprot@cmh.ki.si.

Published November 12, 2021.

*PRACE (<https://prace-ri.eu/>) is an international not-for-profit association with 26 member countries that creates a pan-European supercomputing infrastructure, providing access to computing and data management resources and services for large-scale scientific and engineering applications at the highest performance level.

[†]See the timeline at <https://twitter.com/EricTopol/status/1332771238771630080>.

- At the molecular level, researchers started running molecular simulations of these structures immediately after the basic structural data became available.
- At the global level, epidemiological models were tested and refined, the spread of the virus was modeled, and research groups widely shared these data.

Recognizing the fundamental importance of computer-based approaches to these urgent problems, the European supercomputing centers have established urgent computer access through PRACE, putting into practice for the first time a long-planned idea of a fast-track program for crisis situations. A small scientific committee, which leveraged expertise from the existing PRACE Scientific Steering Committee and PRACE Access Committee, began work on 21 March 2020. The first requests for computing time arrived only a couple of days later. The application system was simple and swift to attract the largest possible number of projects. Fig. 1 shows a schematic depiction of the process. The main requirements for applicants to be awarded were a convincing argument that the proposed research would help with the COVID-19 crisis and a commitment to making the resulting scientific knowledge available to the wider scientific community (thanks to the resources of Fenix ICEI, BioExcel, and EMBL-EBI). On the side of the scientific committee, PRACE's goal was to implement a rigorous and rapid peer-review process.

The scientific committee originally highlighted five research areas as targets for the urgent computing Fast Track Call:

- Biomolecular research to understand the mechanisms of viral infection
- Bioinformatics to understand mutations and evolution
- Biosimulations to develop therapeutics and/or vaccines
- Epidemiological analyses to understand and predict the spread of disease
- A final broad category to understand and mitigate the impact of the pandemic

The scientific committee welcomed any academic or industrial project that included a commitment to open science and open data.

To accelerate collaborative progress based on the outcome of the Fast Track projects, PRACE and its American counterpart, XSEDE (The Extreme Science and Engineering Discovery Environment), jointly launched an international COVID-19 High Performance Computing (HPC) Knowledge Exchange, a fortnightly meeting to foster communication, collaboration, and sharing of code and data. The webinars proved very successful with an average of 50 participants and covered topics such as virtual

screening, drug design, biomolecular simulations, bioinformatics, epidemiological studies, and the spread of virus via droplets.

Fast Track Call Project Assessment

A project in a biannual PRACE Project Access Call needs the positive scientific evaluation of the PRACE Access Committee, plus the complementary technical analysis of the computing centers. Minimum requirements for PRACE Project Access are at least 35 million core hours running on about 1,024 cores concurrently, with good scalability of code, typically 2 Gb per core memory usage, and from 9 to 100 TB of disk space (small variations occur for the different centers).

On the other hand, in the Fast Track Call, science was the main driver and so the scientific committee adapted the requirements regarding the scalability of the code on a case-by-case basis. They customized the requirements regarding the size of the allocation not only by balancing between the large centers but also by bringing in several national institutes.

For PRACE Project Access Calls, the standard response time from the PRACE Access Committee is 5 mo. In stark contrast to this, a much smaller scientific committee drove the COVID-19 Fast Track Call, which mobilized human resources in a variety of settings: Reviewers completed their tasks in a few days, PRACE staff managed the administrative process with high priority, and the supercomputer center staff that provided computing time and storage dealt with allocations in about a week on average.

That the Fast Track Call received more than 50 proposals in just 2 mo documents the determination of the scientific community to help. Over a 4-mo period, PRACE received 80 applications from multiple fields, ranging from the most predictive docking and screening runs for pharmaceutical applications to fluid dynamics studies of droplet dispersion at the physical level to epidemiology on a global scale (the societal level). For comparison, the regular 19th and 20th PRACE calls for proposals for Project Access received 57 and 59 proposals, respectively, from a wide range of scientific disciplines. Fig. 2 provides an overview of the number of proposals (accepted fraction in darker color, rejected one in lighter color, and the total number of proposals received is inscribed in the circle), the domains, and readiness for high-performance computing are shown in Fig. 2. A statistical breakdown of the proposals (country of the principal investigator [PI], sex ratio of the PI, and the number of reviewers per project, including the number of projects in each category and career stage of the PI) is shown in Fig. 3.

The Fast Track Call awarded half a billion computer hours. Successful projects used those resources over a 6-mo period. A typical biannual PRACE Project Access Call awards 2.5 billion computer hours to projects that run for 12 mo. Thus, the Fast Track Call, over a 6-mo period, is equivalent to about 40% of a single PRACE Project Access Call.

Summary of Received Proposals for Each Line of Research

The incoming project proposals quickly showed a broader range of possibilities of COVID-19-related science than the initially proposed research directions (see above), which described mainly biological aspects (mechanisms of viral infection, viral evolution, and drug targeting) and epidemiological elements.

The awarded projects can be roughly divided again into five groups, which the sections below describe in more detail. Out of all the projects submitted, a dozen employ AI techniques,

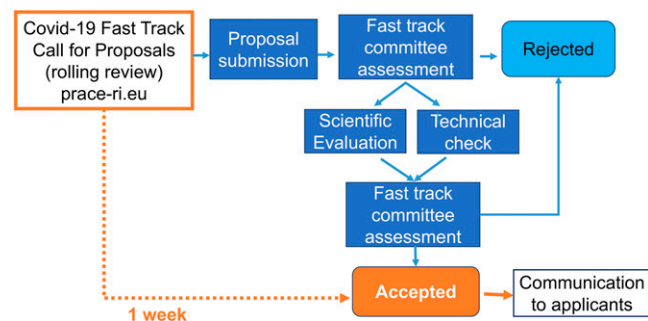


Fig. 1. Scheme of the workflow for applications to the PRACE Fast Track Call for Proposals to Mitigate the Impact of the COVID-19 Pandemic.

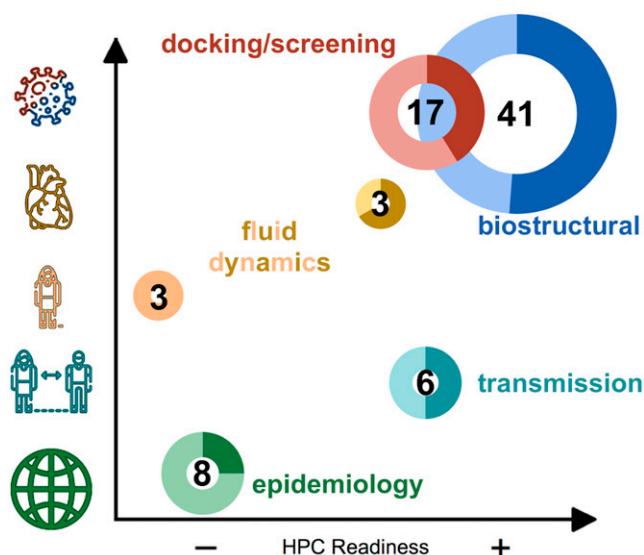


Fig. 2. Overview of 78 of the 80 projects submitted to the PRACE COVID-19 Fast Track Call. Two projects have been omitted as they are beyond the scope. The area of each circle is proportional to the total number of applications received in a given area, as inscribed in the center of the circle. The proportion of projects accepted is highlighted in darker color. The x axis represents the readiness for high-performance computing, while the y axis represents the scale of the objects targeted by each model. The scientific approach is indicated next to each circle: biostructural studies, docking and screening, fluid dynamics at the level of an organ or individual (airborne) transmission, and epidemiology at the global level.

mainly for biostructural research but also for docking/screening aspects.

Biomolecular Research. The biological aspects focused on two main areas of research: a structural one concerned with aspects of the spike protein in the outer wall of the virus (ACE2 receptor) and a pharmaceutical one aimed at drug development. Structural modeling gave the best results in the very porous region of glycan shields: Experiments cannot fully identify these regions, making computational analysis crucial. The numerous biostructural applications focus on molecular mechanisms related to biological function, such as recognition, inhibition, binding, or fusion. They aim at characterizing interactions between relevant macromolecules, their dynamics and conformational changes, and their binding pockets.

Docking/Screening. A subfield related to drug discovery is the (virtual) screening of compounds from existing databases, either to identify new potent compounds associated with a particular disease or to repurpose drugs that may be effective against new targets not previously considered. Researchers in this subfield must take special care, as docking and screening approaches usually yield (many) hits due to the intrinsic stickiness of molecules, even for compounds that later turn out to be irrelevant for medical purposes. Only subsequent experimental (or to a lesser extent independent computational) testing allows an assessment of their eventual relevance. Initially, researchers performed docking runs on the ACE2 spike sites (protein only), but as the structural models developed and improved further they needed to include the shielding glycans (2) surrounding the spike to find realistic drugs for these receptors, and a feedback loop between the structural features of the virus and drug development would

increase the number of successful hits. This crucial role of shielding sugars will probably be one of the most relevant transformations in virtual drug screening in the coming years. One of the projects aimed at finding an antiviral therapy uses a protein structure prediction platform based on machine-learning techniques to help design peptides. Our Fast Track Call received 17 proposals in this area, several with overlapping strategies and pools of compounds. As the scientific committee explicitly required experimental validation, the number of projects in the screening category that received computational resources under the Fast Track Call was limited to six.

Many initiatives specifically targeting compound screening have meanwhile come to fruition, such as the JEDI billion molecules against COVID-19 challenge or the #EuVsVirus initiative from the European Commission, in which PRACE partners provided Tier-0 and Tier-1 HPC resources and user support services during a hackathon. The European Union (EU) reported some progress in this direction, as noted by the Exscalate4CoV consortium. Japan installed its most powerful supercomputer, Fukagu, during the pandemic, and the United States also extensively used HPC resources for COVID-19 docking.

Fluid Dynamics. The initial lines of research neglected the possible areas of organ simulations. However, case studies very quickly showed that the disease affects certain organs in particular, and three projects addressing these aspects received an allocation under the Fast Track Call. Simulations of the heart, lungs, associated organs, and their coupling, are of great importance for safety pharmacology in testing new drugs but also for diagnosis and treatment of the disease, which may involve assisted ventilation. The computer codes used to simulate biofluids, fluid–structure interactions, and their effects on normal organ function were capable of running on supercomputers (i.e., they met scalability requirements). Another category grouped projects that dealt with individual-level studies. These included AI use of data of medical origin and the evolution of viruses by mutation. None of the codes for these computations, or the data the researchers fed them, were suitable for use on HPC.

Transmission. Several projects studied virus transmission via airborne liquid droplets. This process is particularly important in enclosed spaces, such as offices, public transportation, and airplanes, where the directions and positions of ventilation/air conditioning systems can be adjusted to regulate droplet flow to minimize the effects of virus spread. The engineering community uses computational fluid dynamics approaches for this type of research. The associated codes are very ready for HPC applications in terms of scalability and robustness, and three such projects received resources under the Fast Track Call.

Epidemiology. Finally, in epidemiology, three projects consider machine-learning techniques to model the spread of infections, while the others rely on agent-based simulations and compartmental models such as SIR (Susceptible, Infectious, or Recovered, first introduced in 1927) or SEIR (which contains a new class, “exposed”: infected but not infectious) (3). Such models are also useful to facilitate lockdown measures if coupled with population mobility and economics (4), but PRACE did not receive a project proposal that goes in this direction, although this type of project has subsequently been successful in other programs. In this area, most codes were not explicitly HPC-enabled but were still successful in the Fast Track Call. In the context of disease diagnosis and prediction, one project uses

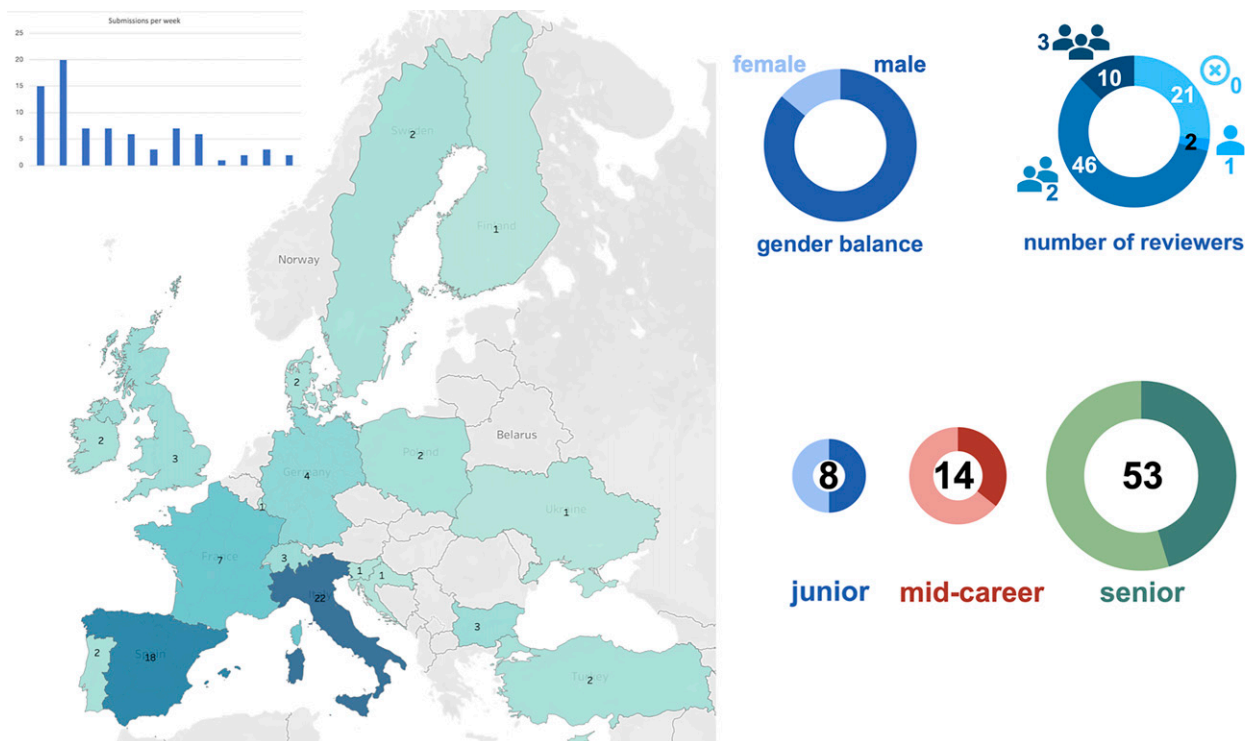


Fig. 3. A statistical overview of the proposals submitted to the Fast Track Call. The map shows the distribution of proposals per country, with an inset of the number of submissions per week since the call was opened. A large number of European countries have participated in the Fast Track Call. The right part of the figure shows the distribution of proposals by sex, the distribution of proposals by career stage, and the number of reviewers per proposal. The number of reviewers subfigure both shows the number of reviewers (outside the circle), from zero to three, and the number of projects for each case (within the circle). In classifying the career stages we have used the categories normally used for the European Research Council (junior: 7 y after PhD; mid-career: 12 y after PhD; senior: more than 12 y after PhD).

machine-learning techniques and computed tomography scans augmented with biomechanical simulations of airflow to train AI models.

Context of the Call and General Observations

The European HPC community has often proposed an EU-wide coordination effort to facilitate access to the next generation of computers through the EuroHPC-JU[‡] research tracks. However, when the pandemic hit Europe, no mechanism had yet been established for rapid and urgent access to massive computing resources on a continental scale. Several national supercomputing centers offered local ad hoc solutions with direct access schemes, without formalized data syndication. The @foldingathome initiative under the Rosetta project (5) is a cloud equivalent to the activities of the supercomputing centers, launched by a self-organized scientific community. If anything, the COVID-19 pandemic has made clear that a mechanism to enable urgent computing with shared open data and fast access is imperative. Nevertheless, simulations are only one pillar of scientific approaches. Mathematical models are instrumental in understanding the pandemic (6) and can predict the occurrence of the next one through surveillance analysis. While they can guide and direct research, uncover new directions, and reduce the number of targets or scenarios to consider, they are even more powerful when integrated into other experimental or technical domains, as mentioned earlier with virtual screening. These computer-based studies often represent the first and essential step in a chain of multidisciplinary, interactive, and iterative

[‡]<https://eurohpc-ju.europa.eu/>.

studies that intertwine experiments and simulations. A fast-track evaluation process can then consider the readiness of subsequent steps.

Lessons Learned. We describe the key lessons learned in the following paragraphs, and these pertain to

- Implementing a rigorous scientific peer evaluation process with well-defined and controlled inherent biases that is resistant to issues such as opportunistic science
- Ensuring availability and openness of produced data and related metadata
- Adapting and possibly relaxing technical requirements to enable all scientific fields to benefit from computational resources
- Encouraging cross-disciplinary proposals with global approaches

Rigorous Scientific Evaluation. To conduct the Fast Track Call, a scientific committee implemented a rigorous selection process for which a high level of reviewer engagement was critical, especially in the special circumstances of a widely locked-down planet. People from all walks of life were PRACE's quick, reliable, critical, and supportive partners. The success rate for the COVID-19 Fast Track was less than 50%, based on scientific and/or technical criteria as well as the potential impact on the pandemic. Given the limited computational resources that were available for the Fast Track Call, the scientific committee decided not to duplicate efforts in any particular direction for thematically overlapping proposals. This strategy favors diversity by exploring a wide range of modalities for crisis management rather than focusing on selected ones. The

scientific committee was aware that the drawback of this could be that preference is given to the first proposals received on a given topic, and not necessarily to the most mature ones. Another problem is opportunistic science (7)—here in the sense of easy access to massive computational resources. The scientific committee filtered out such proposals and sometimes redirected them to PRACE Project Access Calls. For example, docking without external experimental support and testing was rigorously scrutinized, as were medium- to long-term projects unlikely to provide medically tangible data in the short term.

Efficient Sharing of Data. Data curation is an essential tool for ensuring global coordination. It is critical for effectively addressing research challenges such as this pandemic. To ensure that the research community can fully and quickly utilize the data generated by the urgent computations, curated repositories and an open-science approach with associated metadata are crucial. Data must be reanalyzed and made available with all metadata. The data should be documented as thoroughly as possible to enable their wide use. The Research Data Alliance has provided specific COVID-19 recommendations and guidelines.[§] We have found that in data-centric approaches such as AI the data themselves present many challenges. The lack of metadata standards, manual processes, delays in collection, or portability hinder their FAIRness (findability, accessibility, interoperability, and reuse of digital assets), while it is also very difficult to obtain reliable data related to personal behavior or physical medical data, especially for scientists who need to calibrate models and quantify the errors to make informed predictions. In the case of the PRACE Fast Track, PRACE encouraged storing data together with Fenix and provided enough storage space for all the applicants. PRACE called for this data to be disclosed in accordance with the FAIR principles, to ensure that, for example, runs are not repeated and, more generally, that taxpayers' money is spent as transparently as possible.

Enabling All Scientific Fields to Benefit. Is such a call equally beneficial to all areas of science? HPC-ready codes are not available for all fields that are potentially useful in combating the pandemic. Therefore, the scientific committee was eager to support certain proposals, even if they were technically less prepared to run on HPC architectures, as long as they were relevant to advancing the science. These were redirected to national resources. More generally, as computers become faster and more sophisticated, their efficient use is a key concern. To this end, traditional research fields need the support of research software engineers in various areas of scientific computing and visualization. These highly complementary skills should be a constitutive part of any supercomputing infrastructure. Finding a career path for well-trained and versatile software engineers is an urgent need, especially in Europe, as demonstrated by the activities of the High-Level Support Teams in the PRACE resource exploitation process (8).

Encouraging Cross-Disciplinary Proposals with Global Approaches. Applications to the Fast Track Call predominantly targeted well-identified and mostly single-science areas. Ideally, however, more complex and cross-disciplinary models should address the integrated, highly intertwined, and cascading effects that define such a crisis. For example, support for epidemiological models

that incorporate economic concerns would be desirable to evaluate lockdown interventions and achieve a much finer spatial and temporal resolution that limits long-term economic implications. The current emergency movement has highlighted the importance of multilayered, global intervention. Such global approaches should be the goal for future urgent computing implementations. Long-term sustainability of these resources and monitoring of global diseases, natural disasters such as earthquakes, and environmental changes during extreme climate events could be additional areas for further support.

General Considerations, Impact, and Future Urgent Needs

The impact of the PRACE Fast Track Call on pandemic mitigation is still difficult to measure. Even if the first results from the awarded projects are already available and published (see e.g., PRACE Digest 2020, <https://prace-ri.eu/news-media/publications/prace-digest/prace-digest-2020/>), the traction of the initiative might be developing too slowly to meet the urgent need. Fundamentally, this approach is about generating sound scientific knowledge that will form the basis for future rational action and open up new avenues for the development of drugs or practical recommendations and interventions. The time required for such research to be rigorous and robust stands in stark contrast to the urgency of the actions needed. Realistic expectations for such initiatives should be to catalyze and focus research in a particular area, creating a hotspot of scientific activity. This focus should lead to a significant acceleration in selected areas (9)—docking or computational fluid dynamics are good examples in our case—but it cannot, and should not, be expected that any miracle solutions will result from only a few months of calculations. This limitation is especially true because simulations are often only the first, essential step in a chain of value that may lead, for example, in the very last phase to the approval of a new drug. The time needed to implement efficient measures based on such models is therefore usually beyond the short time available to deal with an immediate health crisis. This inherent discrepancy in time scales should in no way diminish the value of these scientific studies and PRACE's efforts to establish urgency calculations in Europe. It should guide the assessment of the results and help to define realistic expectations of such a tool. Interestingly, in our limited experience, areas that would potentially provide meaningful results in a short time, like epidemiology and assessment of nonmedical interventions, were less ready to make optimal use of HPC resources. A more interdisciplinary approach to studying the many aspects of pandemics could be beneficial in creating a monitoring community that is better prepared for emergencies such as the current one. Crucially, this preparation cannot be postponed until the outbreak of a crisis: It must also take place and be supported in "normal" times. More generally, other global challenges, such as climate change, could also benefit from a collective supportive effort. Societies and policy-makers should be empowered to make informed decisions based on statistics, uncertainties, and error estimates, especially in the early stages of a crisis when available information is biased or very limited. Such mathematical skills are crucial for living and thriving in our highly interconnected, very complex modern societies.

Data Availability. Data are available from Zenodo (DOI: [10.5281/zenodo.5607656](https://doi.org/10.5281/zenodo.5607656)).

[§] <https://www.rd-alliance.org/groups/rda-COVID19>.

Acknowledgments

We thank the reviewers for their quick response and rigorous scientific evaluation of the projects. Their tireless efforts were crucial to the development of the Fast Track Call. We also thank the supercomputing center staff for their prompt

assistance in allocating resources and addressing the needs of all successful applicants. Finally, we thank Marjolein Oorsprong for critical reading and copyediting of the manuscript. Icons in Figure 2 were made by Freepik, Smashicons, and Vitaly Gorbachev from <https://www.flaticon.com/>.

- 1 J. Pallesen et al., Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E7348–E7357 (2017).
- 2 B. Turoňová et al., In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science* **370**, 203–208 (2020).
- 3 W. O. Kermack, A. G. McKendrick, Contributions to the mathematical theory of epidemics—I. 1927. *Bull. Math. Biol.* **53**, 33–55 (1991).
- 4 J. M. Brauner et al., Inferring the effectiveness of government interventions against COVID-19. *Science* **371**, eabd9338 (2021).
- 5 M. I. Zimmerman et al., SARS-CoV-2 simulations go exascale to capture spike opening and reveal cryptic pockets across the proteome. *bioRxiv* [Preprint] (2020). <https://doi.org/10.1101/2020.06.27.175430> (Accessed 21 December 2020).
- 6 A. P. Dobson et al., Ecology and economics for pandemic prevention. *Science* **369**, 379–381 (2020).
- 7 H. Else, How a torrent of COVID science changed research publishing - in seven charts. *Nature* **588**, 553–553 (2020).
- 8 E. Lindahl, Ed., “The scientific case for computing in Europe 2018-2026.” <https://prace-ri.eu/about/scientific-case/> (Insight Publishers, Bristol, UK), pp. 6 and 57.
- 9 C. Arnold, How computational immunology changed the face of COVID-19 vaccine development. *Nat. Med.*, 10.1038/d41591-020-00027-9 (2020).