MEDIATE - Molecular DockIng at homE: Turning collaborative simulations into therapeutic solutions

Giulio Vistoli, Candida Manelfi, Carmine Talarico, Anna Fava, Arieh Warshel, Igor V. Tetko, Rossen Apostolov, Yang Ye, Chiara Latini, Federico Ficarelli, Gianluca Palermo, Davide Gadioli, Emanuele Vitali, Gaetano Varriale, Vincenzo Pisapia, Marco Scaturro, Silvano Coletti, Daniele Gregori, Daniel Gruffat, Edgardo Leija, Sam Hessenauer, Alberto Delbianco, Marcello Allegretti & Andrea R. Beccari

Published online: 10 Jul 2023.

Submit your article to this journal ↗

Article views: 732

View related articles ↗

View Crossmark data ↗

Taylor & Francis
Taylor & Francis Group

TECHNOLOGY EVALUATION

🔓 OPEN ACCESS   ⬤ Check for updates

MEDIATE - Molecular DockIng at homE: Turning collaborative simulations into therapeutic solutions

Giulio Vistoli[a]*, Candida Manelfi[b]*, Carmine Talarico[b], Anna Fava[b], Arieh Warshel[c], Igor V. Tetko[d,p], Rossen Apostolov[e], Yang Ye[f], Chiara Latini[g], Federico Ficarelli[g], Gianluca Palermo ⬤[h], Davide Gadioli[h], Emanuele Vitali[h], Gaetano Varriale[i], Vincenzo Pisapia[i], Marco Scaturro[j], Silvano Coletti[k], Daniele Gregori[l], Daniel Gruffat[m], Edgardo Leija[m], Sam Hessenauer[m], Alberto Delbianco[n], Marcello Allegretti[o] and Andrea R. Beccari[b]

ªDipartimento di Scienze Farmaceutiche, Università Degli Studi di Milano, Milano, Italy; ᵇEXSCALATE, Dompé Farmaceutici S.p.A, Napoli, Italy; ᶜDepartment of Chemistry, University of Southern California, Los Angeles, USA; ᵈBIGCHEM GmbH, Valerystr, Germany; ᵉPDC Center For High Performance Computing, KTH Royal Institute of Technology, Stockholm, Sweden; ᶠNatural Products Chemistry Research Center, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, China; ᵍHigh Performance Computing Dept, CINECA, Casalecchio di Reno, Bologna, Italy; ʰDEIB – Politecnico di Milano, Milano, Italy; ⁱSAS Institute Srl, Roma, Italy; ʲSAS Institute Srl, Milano, Italy; ᵏChelonia SA (AG), Allschwil, Switzerland; ˡE4 Computer Engineering S.p.A, Scandiano (RE), Italy; ᵐNanome Inc, San Diego, CA, USA; ⁿEnl, San Donato Milanese, Italy; ᵒDompé Farmaceutici S.p.A, L'Aquila, Italy; ᵖInstitute of Structural Biology, Molecular Targets and Therapeutics Center, Helmholtz Munich-Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH), Neuherberg, Germany

ABSTRACT

Introduction: Collaborative computing has attracted great interest in the possibility of joining the efforts of researchers worldwide. Its relevance has further increased during the pandemic crisis since it allows for the strengthening of scientific collaborations while avoiding physical interactions. Thus, the E4C consortium presents the MEDIATE initiative which invited researchers to contribute via their virtual screening simulations that will be combined with AI-based consensus approaches to provide robust and method-independent predictions. The best compounds will be tested, and the biological results will be shared with the scientific community.

Areas covered: In this paper, the MEDIATE initiative is described. This shares compounds' libraries and protein structures prepared to perform standardized virtual screenings. Preliminary analyses are also reported which provide encouraging results emphasizing the MEDIATE initiative's capacity to identify active compounds.

Expert opinion: Structure-based virtual screening is well-suited for collaborative projects provided that the participating researchers work on the same input file. Until now, such a strategy was rarely pursued and most initiatives in the field were organized as challenges. The MEDIATE platform is focused on SARS-CoV-2 targets but can be seen as a prototype which can be utilized to perform collaborative virtual screening campaigns in any therapeutic field by sharing the appropriate input files.

1. Introduction

Collaborative computing is a consolidated strategy which allows scientific groups distributed worldwide to jointly cooperate to a scientific project by providing their specific contributions without requiring physical interactions [1]. The benefits of remote cooperation to reduce the geographical distances have long been known but became very relevant during the recent Covid pandemic crisis when very tight collaborations could continue while preserving all safety rules [2].

Collaborative computing benefits from the open science paradigm according to which all obtained data along with the utilized methods are freely shared by using collaborative and public platforms [3]. Overall, such an open science paradigm enhances the global scientific capacities by increasing the wealth of available data which the collaborating scientists can explore. The relevance of sharing the scientific data and tools is incredibly enhanced during the recent health crisis [4]. Indeed, the urgent need for efficient therapies to treat the COVID-19 patients is imposing emergency research strategies which must be primarily based on a rapid dissemination of all produced scientific data [5]. Hence, the huge number of open data repositories, dedicated to the multidisciplinary aspects of SARS-CoV-2 and developed in the last few months, comes as no surprise: they include and share epidemiological data often coming from national health agencies, clinical data, economic and social data dealing with the pandemic impacts or omics data mostly focused on viral sequencing [6–9]. Regarding the repositories of drug discovery data, many purposely developed resources share data from

---

Article highlights

- The COVID-19 pandemic crisis prompted a tremendous effort in computational analyses for drug repurposing although the reported virtual screening campaigns cannot be synergistically combined.
- Docking simulations and virtual screening campaigns might be easily combined in collaborative projects if all involved researchers work on the same input files while applying their preferred computational procedures.
- MEDIATE initiative, developed by the E4C consortium, is based on a platform specifically developed to support COVID-19 drug repurposing in which prepared compounds' libraries and protein structures are shared to enable standardized virtual screening campaigns.
- The contributions (docking results) submitted to MEDIATE will be utilized to develop global consensus models by AI techniques and the most promising compounds will be experimentally tested.
- While being targeted on COVID-19, MEDIATE can represent a prototype of collaborative platform for performing standardized virtual screening campaigns by sharing the required input structures.

computational studies as well as biological data mostly generated by HTS screening based on drug repurposing campaigns [10]. In this context, the Exscalate4CoV (E4C) consortium is committed to rapidly and publicly disseminate all generated results and such a commitment is further empowered by the Manifesto recently launched by the European Commission to maximize the accessibility of research results to combat the SARS-CoV-2 pandemic crisis [11].

The SARS-CoV-2 pandemic prompted a never seen effort in computational analyses focused on virtual screening and drug repurposing. A simple literature search on PubMed (accessed on 13 January 2023) using the keywords 'covid virtual screening' returned 1795 publications and the more generic terms 'covid docking' provided 3740 results. Notably, 2021 and 2022 show in both cases roughly the same number of publications suggesting that such a huge computational effort is a still ongoing process. Although many of these repurposing studies involved rather similar datasets of safe-in-man molecules and the number of therapeutically relevant viral proteins is quite limited, the results of these studies can hardly be combined. The reasons for this difficulty are varied and include, above all, the differences in both ligands' datasets (and ligands' set-up) or in the simulated protein structures. Not to mention that, even though the results are freely shared, the raw data of the docking simulations are rarely available. While appreciating the scientific richness of these publications, the impossibility of synergistically combining their results is clearly pithy, especially considering that a consensus combination could easily be achieved if the Authors share ligands' datasets and protein targets while applying their preferred computational strategies. Stated differently, structure-based virtual screening campaigns are well suited for collaborative computing provided that the involved groups work on the same input files [12,13]. The efficiency of such a collaborative consensus strategy depends on the number of considered targets since a high number of simulated proteins might provide too dispersive results which cannot be efficiently combined. In contrast, antiviral campaigns mostly focused on the viral targets (as in the here reported case for SARS-CoV-2) could be particularly

suitable since they comprise a limited number of proteins and the consensus of the submitted docking results should support the identification of potent antiviral compounds.

On these grounds, the E4C consortium presents here the MEDIATE initiative (MolEcular DockIng AT home, https://mediate.exscalate4cov.eu accessed on 17 January 2023) which combines the richness of the modeling data generated and shared by the consortium with the powerfulness of an environment specifically designed to organize and exploit the efforts from collaborative computing. As depicted in Figure 1B and by means of the MEDIATE initiative, the E4C Consortium shares a set of prepared and standardized input files (including refined ligands and annotated targets) and collects the docking results that groups worldwide generate by using these input files and applying their preferred computational procedures. Hence, this initiative aims to integrate the predictive analyses performed by different scientific groups to arrive to a sort of global consensus, based on Machine Learning (ML) and Artificial Intelligence (AI) technologies, to assure robust and method-independent predictions. The best compounds selected by such a worldwide consensus strategy will be then purchased and tested. The biological results will be finally published and shared with the scientific community.

2. Overview of the market

The recent pandemic crisis has led to a significant impact on the open science paradigm and several initiatives appeared in the last years. As recently reviewed, these initiatives can be subdivided into two groups [14]. On one hand, the pandemic fostered the wide dissemination of the obtained results by supporting both open access publications and open data sharing. With regard to open access, the pandemic effect is clearly documented by the remarkable growth of the preprint servers: the number of published preprints is almost tripled in 2020 and 2021 compared to the previous years. Also, the open data initiatives enjoyed a similar increase with dedicated resources such as the NCATS OpenData COVID-19 (https://opendata.ncats.nih.gov/covid19 accessed on 17 January 2023) which is mostly focused on the efforts in drug repurposing. On the other hand, the pandemy promoted open science collaborative projects among which the here described MEDIATE initiative can be included. Among the other open projects developed during the pandemic, one may mention, for example, the PostEra COVID Moonshot [15] which invites the scientists worldwide to submit their ideas for antiviral compounds which are evaluated and prioritized by the PostEra AI technologies. The most interesting compounds are then synthesized and experimentally tested against the SARS-CoV-2 main protease. All the obtained results are made publicly available. Another example is represented by the JEDI GrandChallenge which is organized in three phases [16]. The first step invites the research teams around the world to virtually screen billions of compounds against the viral targets. The submitted results will be compared and cross-related to extract a list of highly promising molecules which will undergo to in vitro (phase 2) and in vivo (phase 3) testing. The phase 1

Figure 1. The MEDIATE initiative: (a) Graphical overview of the contributions of the groups from different countries participating in the initiative and (b) Main actions involved in the initiative leading to testing and sharing identified antiviral compounds.

is now finished reaching 54 billion molecules virtually screened by 130 teams. Teams which proposed the most active molecules will be awarded. Other open projects involve the free distribution of compounds which possess antiviral activity for further investigations (MMV COVID Box) [17], while Folding@home (F@H) [18] exploits its distributed computing platform to investigate the interactions between the SARS-CoV-2 Spike protein and the human ACE2 receptor.

In addition to the COVID-inspired initiatives, the SAMPL (https://www.samplchallenges.org, accessed on 16 January 2023) e D3R (https://drugdesigndata.org, accessed on 16 January 2023)

challenges represent relevant and long-lived examples of collaborative calculations in drug discovery (not only focused on docking simulations) in which the competitive aspect played a key role. In detail, D3R organized five docking challenges starting from 2015 and for each challenge the required goals involved the prediction of both the crystallographic poses and the affinity rank for a set of selected ligands. For each challenge, various targets were proposed including, among others, Cathepsin S, Farnesoid X receptor (FXR), BACE, and HSP90. Notably, datasets and submitted results for each challenge are still available and all scripts utilized to analyze the docking results are available on

GitHub. Each challenge resulted in an overview publication plus a special issue collecting papers written by participants about their methods and results [19]. These papers represent a remarkable benchmarking review of the best performing approaches for both docking simulations and free energy calculations. The SAMPL challenges were more focused on the prediction of physicochemical properties such as log P, pKa, and solubility. Nevertheless, some SAMPL challenges were devoted to ligand docking as well as to the host-guest binding process: for example, the SAMPL7 challenge involved the docking of a representative set of fragments to the Pleckstrin homology domain interacting protein (PHIP) [20], while SAMPL5 comprised a challenge based on the aqueous host-guest binding processes for different host molecules [21]. In this context, the Critical Assessment of protein Structure Prediction (CASP) experiment (https://predictioncenter. org/, accessed on 1 April 2023) is focused on the modeling of protein structures. This invites scientific groups worldwide to predict the structures of proteins that are about to be experimentally resolved. The analysis of the submitted predictions allows a critical benchmarking of the available methods (often implemented in free web services), of the advancements made in the field and of the most promising future directions.

3. The MEDIATE collaborative platform

3.1. Input files shared by MEDIATE

Protected by a secure account system, MEDIATE shares several input files which include both compounds' libraries and protein structures specifically prepared and annotated to support optimized and standardized virtual screening campaigns. The complete list of the downloadable libraries is reported in Table 1. For each dataset, the following properties are used to characterize the molecules: Molecular weight (MW); the octanol/water partition coefficient (Log P) calculated using the atom-based method published by Ghose and Crippen [22]; Polar Surface Area (PSA) calculated using a 2D approximation [23] and the number of rotatable bonds (N rotors).

All compounds were converted to 3D and prepared with Schrödinger's LigPrep tool (Schrödinger Release 2020–2: LigPrep, Schrödinger, LLC, New York, NY, 2020.). This process generated multiple states for stereoisomers, tautomers, ring conformations (one stable ring conformer by default), and protonation states. The Schrödinger package Epik was used to assign tautomers and protonation states that would be

dominant at a selected pH range (pH = 7 ± 1). Ambiguous chiral centers were enumerated, allowing a maximum of 32 isomers to be produced from each input structure. Then, the energy minimization was performed by using the OPLS3 force field [24]. Duplicates among the libraries were removed.

The shared libraries allow various virtual screening campaigns to be performed since they comprise (1) commercially available molecules for hit identification, (2) drugs, natural and food compounds for repurposing studies, and (3) di, tri, tetra, and pentapeptides for the rational design of peptide binders. In particular, the 'Drugs' library includes the set of safe-in-man drugs, either marketed or under active development in clinical phases.

While a detailed analysis of the chemical space covered by the proposed libraries goes beyond the scope of this paper, the average property values for the commercial compounds are in agreement with the values reported by recent comparative analyses based on the chemical space of currently purchasable compounds libraries [25]. Gratifyingly, drugs and natural compounds are in line with the property averages seen for commercial compounds, while foods include on average larger and hugely more lipophilic compounds reasonably due to the occurrence of fatty acids and other lipid derivatives. As expected, peptides appear to be larger, more polar and more flexible molecules compared to the other compounds.

Concerning the protein structures, MEDIATE is primarily focused on the viral targets (apart from ACE2 and TMPRSS2 proteins). As stated in the Introduction, this choice limits the number of simulated proteins and allows the generation of docking results which should be successfully combined. In detail, the number of involved targets comprises 14 proteins including both viral (12) and host (2) structures. The complete list of the available targets is reported in Table 2.

For each target, MEDIATE provides at least one prepared protein structure with at least one annotated binding pocket. Moreover, for six relevant structures, more than one protein structure is stored since they are representative of different conformational states of the protein and thus can be useful for ensemble docking experiments. The shared protein structures are constantly updated based on the availability of new relevant resolved structures. Amid the protein structures there are 12 experimentally resolved structures and two theoretical models. Overall, the available protein structures comprise 40 annotated pockets which belong to 18 different binding sites, including both orthosteric and allosteric annotated pockets. In

Table 1. Shared compounds' libraries available within the MEDIATE initiative with the resulting average values (± standard deviations) for the monitored properties. All the reported properties are computed by Pipeline Pilot.

Library	n. molecules	MW (g/mol)	Log P	PSA (Å²)	N rotors
Commercial Compounds MW < 330 (C-LMW)	1,899,269	261.17 ± 52.14	1.87 ± 1.34	61.54 ± 25.55	3.65 ± 1.79
Commercial Compounds 330<MW < 500 (C-MMW)	2,815,278	400.8 ± 47.29	3.1 ± 1.5	88.1 ± 27.61	5.71 ± 2.11
Commercial Compounds MW > 500 (C-HMW)	249,982	554.58 ± 59.84	4.87 ± 2.01	110.27 ± 40.65	8.35 ± 3.52
Commercial Compounds overall averages	4,964,529	355.13 ± 49.78	2.72 ± 1.64	79.06 ± 27.48	5.05 ± 2.06
Natural compounds (NC)	263,529	435.33 ± 186.15	3.10 ± 2.89	101.77 ± 76.82	6.35 ± 6.1
Drugs (DRG)	8,721	414.84 ± 183.4	2.55 ± 2.94	101.81 ± 76.04	6.87 ± 5.66
Foods (FOOD)	65,461	722.39 ± 262.73	14.62 ± 8.59	93.31 ± 63.66	37.91 ± 20.45
Dipeptides (2PEP)	400	296.84 ± 42.6	−1.50 ± 1.33	142.46 ± 27.18	7.60 ± 1.77
Tripetides (3PEP)	8,000	415.72 ± 52.10	−1.87 ± 1.64	192.14 ± 33.25	11.40 ± 2.16
Tetrapeptides (4PEP)	160,000	534.61 ± 60.16	−2.25 ± 1.92	241.82 ± 38.39	15.2 ± 2.51
Pentapeptides (5PEP)	3,200,000	653.49 ± 67.26	−2.64 ± 2.18	291.50 ± 42.92	19 ± 2.79

Table 2. Prepared target structures with annotated binding sites downloadable from the MEDIATE initiative. The target structures and the annotated binding pockets are those reported in [26] apart from the NSP-13 target which was updated by adding a recently resolved structure (PDB Id: 6XEZ) [27]. The list also includes the TMPRSS2 structure and relative binding sites by using the resolved structure PDB Id: 7MEQ [28].

	Source	Structures	Orthosteric sites	Allosteric sites
3CL-Pro	X-ray	17	14	3
N-term	X-ray	2	2	0
NSP-3	X-ray	1	1	0
NSP-6	Model	2	2	0
NSP9	X-ray	1	1	0
NSP-12-NSP-7-NSP-8	X-ray	1	1	2
NSP-13	X-ray	1	1	1
NSP14–10	Model	1	1	0
NSP-15	X-ray	2	2	0
NSP16–10	X-ray	2	2	0
PL-Pro	X-ray	1	1	0
Spike	X-ray	4	4	0
ACE2	X-ray	1	1	0
TMPRSS2	X-ray	1	1	0

detail, the downloadable target structures define 14 orthosteric and 4 allosteric binding sites. Each pocket is described by a set of structural and physicochemical features as computed by FPocket [29] plus some geometrical data such as the coordinates of the center of the pocket and its size in terms of both radius of the sphere and sides of the box encompassing the cavity. These data are essential to perform standardized docking simulations and are generated by the Pocket program [26] implemented by the VEGA software [30].

To enrich the quantity and quality of the structural data of the viral proteins, we also contributed to a platform called SCoV2-MD (www.scov2-md.org, accessed on 17 January 2023) that systematically organizes atomistic simulations of the SARS-CoV-2 structural and non-structural proteins. The database includes simulations produced by leading groups using molecular dynamics methods to go through the structure-dynamics-function relationships of viral proteins [31]

3.2. Expected results from scientific community

The founding idea of the MEDIATE initiative is that each scientific group involved in virtual screening campaigns has developed during its research activities a set of computational workflows that provide optimized performances in well-defined conditions. Even though all these computational strategies are published, we believe that only the researchers who developed them have the required sensitivity to apply them as best as possible to maximize the resulting performances. Hence, we invite the interested researchers to apply their preferred computational strategies to perform structure-based virtual screening campaigns by using the shared compounds libraries and the prepared protein structures. As described below, the submitted results will be combined to select the most promising compounds to be experimentally screened. The results of docking simulations can be submitted to the MEDIATE initiative by uploading a text file including three fields. In detail and for each docked compound, the file must include: (1) the docking score (that will be then normalized accordingly to the software used), (2) the compound ID, and (3) the binding site ID. The coordinates of the computed complexes are not required. Note that all compounds included in the shared libraries are defined by a unique ID number and

the same holds true for the 40 annotated binding pockets. To complete the submission of the results, the user must provide information regarding: (1) the used docking program, (2) the architecture of the exploited computational resources and (3) whether the simulation was supported by a supercomputer center or was funded by specific grants.

Thanks to the collaboration with SAS (https://www.sas.com/en_us/home.html, accessed on 19 January 2023), the collected data will be used to generate predictive models using the most advanced Machine Learning and AI techniques, which will allow the development of a global ranking to select, among all the ligands, the best candidate molecules. Moreover, the consensus methods are focused both on a single protein to find the most promising molecules for each target and on cross-target methods (with a polypharmacological approach) to find potentially active ligands on the greatest number of targets. The selected molecules will then be purchased and/or synthesized and submitted for experimental validation.

3.3. Experimental testing

The compounds, which will be selected by combining all the docking results using AI methods, will be tested in the following assays for screening of activity against SARS-CoV-2:

- Evaluation of the antiviral activity of SARS-CoV-2 in cell-based assays by HTS. The first is a cell-based assay on green-monkey VERO-E6 cells which constitutively expresses EGFP fluorescent protein. This cell line has been extensively used for SARS-CoV-like viruses' studies and is highly susceptible to cell death after infection [32]. Thus, cell growth is a commonly accepted parameter to monitor cytotoxicity induced by viral infection. Through fluorescence signal quantification, we easily follow cell viability (reported as % of Confluence in the table). % Confluence is calculated by quantification of the total surface of the field that gives a green fluorescent signal, due to the presence of EGPF-positive cells in SARS-Cov2 infected cells treated with test compounds compared to untreated control wells: higher values mean that there are a large number of cells on the microtiter plate

bottom surface, small values mean that most of the fluorescence is lost (i.e. the cells died).

- Evaluation of biochemical assays to test the following SARS-CoV-2 proteins by the Fraunhofer-Gesellschaf Institute:

  (I) **3CL-protease**. This is a biochemical cell-free assay that aims to evaluate the ability of a compound to interfere with the protease activity of the 3CLpro viral protein. The detection of enzymatic activity of the SARS-CoV-2 3CL-Pro will be performed under conditions like those reported by Zhang et al. [33]. Enzymatic activity will be measured by a Förster resonance energy transfer (FRET), using the dual-labeled substrate, DABCYL-KTSAVLQ↓SGFRKM-EDANS (Bachem #4045664) containing a protease-specific cleavage site after the Gln. In the intact peptide, EDANS fluorescence is quenched by the DABCYL group. Following enzymatic cleavage, generation of the fluorescent product was monitored (Ex/Em = 340/460 nm) (EnVision, Perkin Elmer).

  (II) **PL-protease**. The ability of inhibiting the Papain-like Protease Protein activities will be evaluated on a FRET-based assay. The biochemical assay for the detection of PLpro enzymatic activity was developed in accordance with recent publications by Shin et al. [34]. Here we will use a commercial source of the protein (BPS Bioscience #100735) and a fluorescently labeled ISG15 as a substrate (BostonBiochem ISG15/UCRP AMC, #UL-553). The assay will be performed in a buffer containing 50 mM Tris (pH 7.5) and 150 mM NaCl, using 100 nM of SARS-CoV-2 3CLpro and 2.5 μM FRET-substrate.

  (III) **RNA Helicase. nsp13 unwinding-associated activity**. The assay for detection of SARS-CoV-2 helicase activity was developed based on the FRET assay reported by Adedeji et al., 2012. The assay uses a forked double stranded DNA substrate with a 5'- BHQ-2 quencher on the leader strand and a 3'-Cyanine-3 fluorophore on the second strain. The helicase opens the dsDNA substrate in 5'-3' direction releasing the Cy3-labeled strain, signal is detected at Ex/Em = 531/590 nm using the Perkin-Elmer Envision multimode microplate reader. Rebinding of the quencher-labeled strain is inhibited through the addition of a non-labeled strain representing the binding region of the fluorophore-labeled strain.

  (IV) **Polymerase.** The activity of RdRp is detected based on the increase in fluorescence upon binding of the picogreen stain to dsDNA, dsRNA or DNA:RNA hybrids. dsRNA fragments are generated by an active SARS-CoV-2 RdRp. Picogreen is highly selective for double-strand fragments over single-strand DNA or RNA. The signal is detected at Ex/Em = 485/535 nm using the Perkin-Elmer Envision multimode microplate reader.

4. Applications

To date, 60 laboratories from 16 different countries have joined in the initiative (Figure 1a). Among the groups that already submitted their results, we can mention University of Milan (ITA), University 'Magna Graecia' of Catanzaro (ITA), University 'Federico II' of Napoli (ITA), University of Siena (ITA), University of Tuscia (ITA), Korea Institute of Science and Technology Information (KOR), EGE Üniversitesi (TUR), Åbo Akademi University (FIN), and Universitat Rovira i Virgili (ESP). The scientists involved in this project have carried out their computational studies using specific molecular docking codes on the datasets of ligands and proteins provided by MEDIATE. The results of the simulations were uploaded to the web platform and the data collected so far allowed us to perform useful analyses and to extrapolate some preliminary results.

The results shown in the plot of Figure 2 report the average values of the normalized docking scores of each ligand dataset
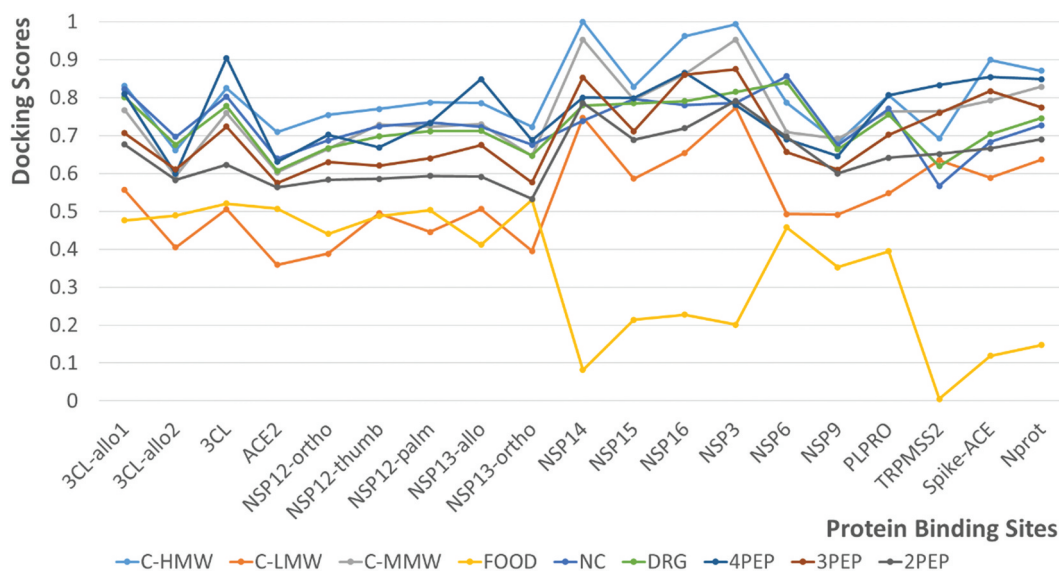

Figure 2. Average normalized docking scores of compound's libraries on the shared proteins (the normalized values increase with the goodness of the complexes).

in relation to the viral and host proteins available on MEDIATE. The data allow some considerations to be drawn. Firstly, the docking scores computed for the various datasets in the 19 explored pockets show comparable profiles. An exception is represented by the trend of the FOOD's scores which is clearly lower and can be related to the physicochemical properties of the molecules. In particular, the high values of MW and rotatable bonds render these molecules poorly suitable for the binding sites of viral proteins.

In more detail, the analyzed docking results unravel characteristic trends that can be related to the structural and physicochemical properties of ligands and pockets. In general, the obtained results confirm that docking algorithms tend to give scores which increase with the molecular weight, even though, in some specific cases, the characteristics of the ligands or classes of ligands may follow opposite trends.

Concerning the binding pockets of 3CL-Pro (Figure 2), the orthosteric pocket shows the highest score averages for all the screened libraries. In detail, the highest values are recorded by tetrapeptides, commercial molecules of high molecular weight and natural products. This confirms

that the orthosteric site prefers larger ligands compared to the two allosteric pockets.

Another important effect is seen for the NSP9 protein. In this case, all the screened libraries show low score averages. The rationale behind this evidence is related to the fact that the pocket of the protein belongs to a particular solvent-exposed region (Figure 3a). Hence, the low number of useful interactions that a ligand can elicit with NSP9 could explain this trend.

An interesting example regards the host protein TRPMSS2. The scores obtained in this pocket are inversely proportional to the size of the ligands for each dataset. In fact, high-MW commercial and natural compounds reveal the lowest docking averages with respect to other protein binding sites. This trend is also closely related to the morphological characteristics of the binding site, which is particularly narrow (Figure 3b).

As proposed in a recent study [35], further analyses were performed to compare the so retrieved best hits. For each dataset and for each target, the top 100 scored molecules were extracted and utilized for pairwise comparisons between targets. For the 'drugs' library, Figure 4 reports
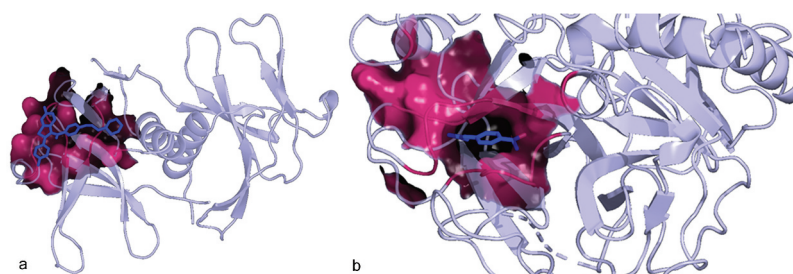

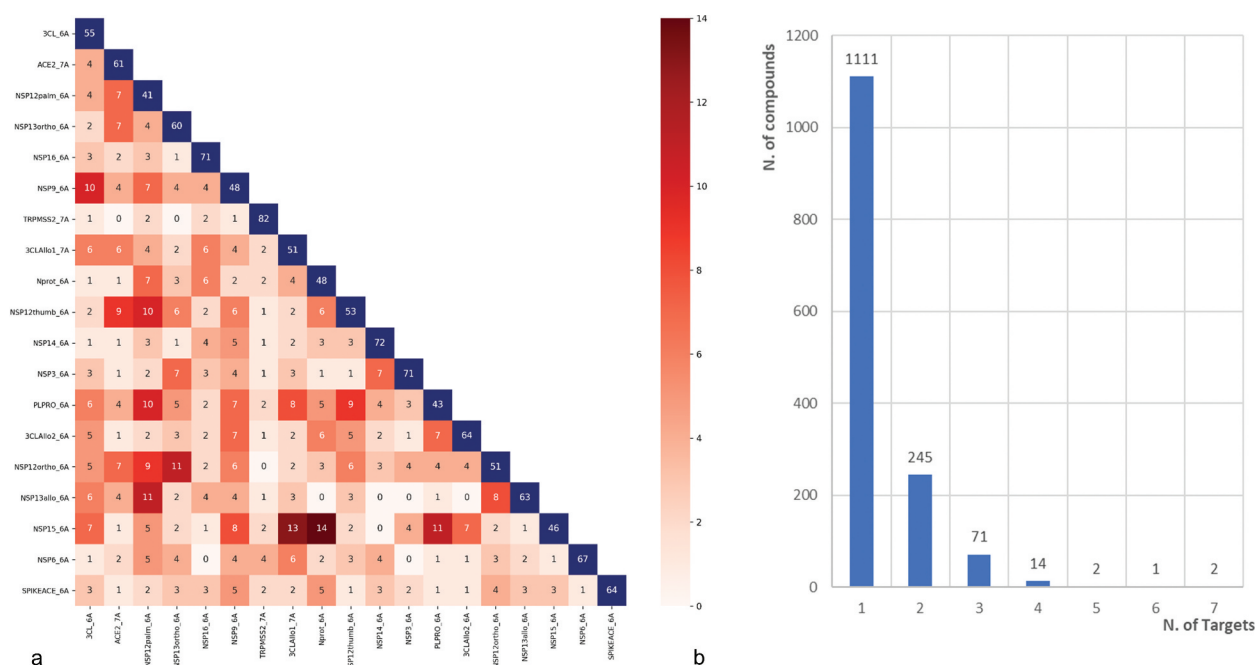Figure 3. Graphical representation of the NSP9 (3a) and of the TRPMSS2 (3b) binding sites.

Figure 4. Number of top scoring molecules common to each pair of binding sites for the "drugs" library (4a). Each off-diagonal box (red shades) contains the number of compounds occurring in the top 100 compounds for both sites. The diagonal boxes (blue shades) report the unique top scoring compounds for each site. Distribution of the number of shared compounds found in the top 100 compounds in respect to the number of the targets (4b).

the number of identical top-scored compounds observed between target pairs with the unique retrieved compounds for each site in the diagonal boxes.

The low numbers of shared top-ranked compounds between target pairs as well as the high numbers of unique ligands emphasize that a few compounds were selected based on criteria mostly depending on the ligand properties and most analyzed scores encode for good interactions in specific binding sites. Even though focused on a reduced portion of the top-ranked molecules (Top 100 vs. Top 500), the here reported shared molecules appear markedly less frequent when compared to the previous study [35].

The distribution of the number of shared compounds versus the number of common targets (Figure 4) confirms that a vast majority of compounds (75%) are predicted to selectively bind to only one target, while some compounds were found to potentially bind to more than one target, up to seven binding sites with a decreasing frequency. Considering that no target pair is markedly richer in shared top-scored compounds compared to all target pairs, we can exclude the presence of particularly promiscuous binding sites that can be targeted by many different compounds. Thus, this approach might prove successful to select molecules with a potential polypharmacological profile on SARS-COV-2 proteins.

Table 3 compiles the correlations between the normalized docking scores of target pairs. The average value of all the correlation values, corresponding to 0.34, confirms that there is not a strong bias due to the ligand effect.

Finally, Table 4 reports some of the top-scored molecules, selected for the best poly-pharmacological profile, that have been experimentally validated or are in clinical trials for COVID-19. For each pocket, Table 4 reports in bold the normalized scores which are higher than the corresponding mean values plus twice the standard deviation, which is a well-known statistical criterion used here as a threshold to identify potentially active molecules. This analysis confirms the goodness of the used docking methods as well as of the statistical parameter applied to filter out the results, considering that all the 10 analyzed drugs show least one and, in some cases, up to 3–4 docking scores higher than the defined threshold.

5. Novelty of the MEDIATE initiative

As stated in the Introduction, virtual screening campaigns can be well suited for collaborative computing projects provided that all involved groups share the same input files concerning both the ligands' datasets and the target structures. Notwithstanding this, such a quite simple concept is rarely exploited to promote collaborative virtual screening campaigns. Indeed, the previous initiatives were primarily based on the challenging concept by which the scientists are left free to choose both the computational strategy and the input files. For example,

the JEDI grand challenge had a primary objective to collect as many docking results as possible in a short time. Hence, JEDI required that each participant submitted a huge number of simulated molecules by three different docking strategies. While suggesting some relevant ligands collections (e.g. the ZINC library), this initiative was less constraining in the selection of the input files. The comparison and consensus of the submitted results were performed by relying on the cross-relations between the top-score compounds submitted by the various groups. Such a strategy provided remarkable results in terms of submitted docking simulations and retrieved active compounds but required massive computational efforts often supported by HPC infrastructures.

Compared to the JEDI challenge, collaborative virtual screening based on shared input files as proposed by the MEDIATE initiative should optimize the potential for success while minimizing the required computational effort and maximizing the overlapping between the submitted results. Moreover, each group can participate in the MEDIATE initiative based on its computing power since MEDIATE shares several ligands datasets of different size and complexity and each participant is free to choose which datasets to simulate. Similarly, MEDIATE shares a total of 37 prepared protein structures including 40 annotated binding sites but does not require that the submitted docking calculations involve all these targets since each group can contribute by selecting targets as they see fit. Such an organization should lead to a better distribution of workloads since each contributor submits its docking results depending on its computational possibilities and the shared input files assure a complete homogeneity and comparability of all submitted simulations. In other words, MEDIATE would benefit from the computational expertise gained by each participant rather than from its computing power.

For the sake of completeness, it should be remembered that other similar initiatives were based on the employment of shared input files but they had different objectives compared to MEDIATE. As described above, the SAMPL e D3R challenges are based on shared input files, but, even when they are focused on docking simulations, they involved a limited number of ligands and the challenges awarded the capacity to predict both their experimental poses and the rank of activity. Basically, these initiatives are based on the capacity to reproduce not yet disclosed experimental data. Strategies based on the contribution of various participating research groups have been successfully applied in computational toxicology to develop consensus models to predict acute toxicity [36] as well as to identify endocrine disrupting chemicals [37]. To the best of our knowledge, MEDIATE is one of the first initiatives in which the advantages of shared input files are exploited to organize collaborative virtual screening campaigns.

Table 3. Docking score correlation between SARS-COV-2 target pairs.

	3CL	ACE2	NSP12palm	NSP13ortho	NSP16	NSP9	TRPMSS2	3CL Allo1	Nprot	NSP12thumb	NSP14	NSP3	PLPRO	3CL Allo2	NSP12ortho	NSP13allo	NSP15	NSP6
ACE2	0.34																	
NSP12palm	0.50	0.36																
NSP13ortho	0.38	0.34	0.39															
NSP16	0.27	0.10	0.28	0.13														
NSP9	0.34	0.26	0.39	0.25	0.47													
TRPMSS2	0.06	−0.11	0.01	−0.12	0.33	0.21												
3CLAllo1	0.47	0.31	0.49	0.34	0.42	0.46	0.13											
Nprot	0.25	0.09	0.23	0.07	0.57	0.52	0.40	0.37										
NSP12thumb	0.44	0.32	0.50	0.34	0.39	0.47	0.07	0.53	0.33									
NSP14	0.18	0.00	0.14	−0.01	0.49	0.47	0.50	0.28	0.58	0.25								
NSP3	0.21	0.06	0.21	0.06	0.53	0.52	0.42	0.35	0.58	0.32	0.63							
PLPRO	0.47	0.32	0.51	0.33	0.37	0.44	0.10	0.51	0.32	0.49	0.23	0.29						
3CLAllo2	0.42	0.31	0.43	0.31	0.35	0.46	0.12	0.45	0.32	0.44	0.26	0.31	0.43					
NSP12ortho	0.38	0.37	0.45	0.37	0.15	0.29	−0.08	0.35	0.11	0.38	0.03	0.08	0.38	0.34				
NSP13allo	0.43	0.30	0.45	0.36	0.31	0.41	0.09	0.45	0.30	0.47	0.23	0.28	0.43	0.40	0.36			
NSP15	0.40	0.23	0.41	0.23	0.56	0.54	0.28	0.56	0.54	0.46	0.45	0.51	0.49	0.41	0.27	0.40		
NSP6	0.47	0.33	0.50	0.32	0.43	0.50	0.13	0.56	0.39	0.53	0.31	0.36	0.48	0.46	0.36	0.46	0.54	
SPIKEACE	0.25	0.09	0.21	0.11	0.52	0.47	0.41	0.33	0.59	0.29	0.54	0.58	0.31	0.30	0.12	0.32	0.51	0.37

Table 4. Normalized score values for some experimentally validated in clinical trials for COVID-19 drugs found among the selected top-scored molecules. The normalized scores which are higher than the corresponding mean values plus twice the standard deviation are highlighted in bold.

Name	3CL	ACE2	NSP12-palm	NSP13-ortho	NSP16	NSP9	TRPMSS2	3CL-allo1	Nprot	NSP12-thumb	NSP14	NSP3	PLPRO	3CL-allo2	NSP12-ortho	NSP13-allo	NSP15	NSP6	Spike-ACE
MONTELUKAST	**0.73**	0.5	0.61	0.49	0.63	**0.67**	0.21	0.62	0.66	**0.64**	**0.78**	0.67	0.54	0.61	0.44	**0.65**	0.67	0.59	**0.7**
TELMISARTAN	0.54	**0.61**	0.47	0.46	0.62	0.55	0.59	0.63	0.52	**0.63**	**0.73**	0.61	0.59	0.58	0.6	0.52	0.58	**0.74**	0.41
BEMCENTINIB	**0.74**	0.43	0.63	0.59	0.56	0.55	0.21	0.55	0.62	**0.64**	0.57	0.53	0.67	**0.64**	0.53	0.61	0.56	**0.75**	0.6
TRADIPANT	0.6	0.44	0.63	0.52	0.59	0.51	0.52	**0.72**	0.6	0.48	0.7	0.61	0.59	0.39	**0.61**	0.6	0.65	0.57	0.56
ATORVASTATIN	**0.85**	0.4	0.51	0.42	0.68	0.59	0.46	0.58	0.55	0.38	0.53	0.67	0.6	0.59	0.5	0.42	0.62	0.64	0.63
BREQUINAR	0.54	0.38	0.5	0.41	0.65	**0.72**	0.5	0.52	0.59	0.62	**0.71**	0.59	0.58	0.56	**0.6**	0.45	0.65	0.46	0.53
CLOFAZIMINE	0.64	0.46	0.54	0.47	0.56	0.44	0.63	0.66	0.58	0.47	0.59	0.53	0.52	0.46	0.37	**0.66**	**0.71**	0.64	0.54
PACRITINIB	0.6	0.51	0.61	0.46	0.51	0.57	0.21	0.54	0.58	0.49	**0.78**	0.52	0.63	**0.63**	0.57	0.49	0.6	0.65	0.51
DUVELISIB	**0.76**	0.4	0.51	0.46	0.69	0.61	0.39	0.66	0.54	0.49	0.58	0.6	0.61	0.51	0.38	0.53	0.61	0.58	0.53
CANDESARTAN	**0.81**	0.41	0.44	0.51	**0.71**	0.59	0.47	0.44	**0.72**	0.38	0.63	0.62	0.36	0.5	0.46	0.46	0.64	0.65	0.43

6. Conclusions

As detailed above, the primary objective of the MEDIATE initiative is to benefit from the worldwide disseminated expertise of the research groups involved in virtual screening studies to discover an optimized set of potential SARS-CoV-2 inhibitors to be tested and shared with the scientific community in reasonable time. The objective is pursued by inviting all interested researchers to contribute with their simulations based on which enhanced and cross-related predictive models and consensus strategies can be suitably developed. To allow a successful combination of the submitted predictions, standardized virtual screening campaigns should be performed. To this end, MEDIATE provides annotated input files (for both proteins and ligands) to be used for this initiative, while the researchers are free to apply their preferred (and purposely optimized) computational procedures. In the context of the E4C consortium, the identified most promising compounds will be experimentally tested and the results will be rapidly published and deployed to the scientific community.

The E4C consortium was involved in a notable experimental effort to validate the results coming from all the computational activities. This allowed the identification of about 600 new active molecules by producing more than 70,000 screening data [38]. These experimental activities involved both biochemical and phenotypic assays. As example of the biochemical studies, the Exscalate4CoV project recently reported an experimental screening study in which about 7,000 safe in human molecules were tested and 105 potent SARS-CoV-2 3CL-Pro inhibitors were identified (with IC50 < 10 μM) [39]. Concerning the cellular analyses, the results from a large-scale repurposing campaign by cytopathic SARS-CoV-2 screening on VERO-E6 cells have been recently published [40]. Based on all these results, focused datasets were subsequently analyzed, and novel antiviral compounds were identified by an iterative screening approach in which computational and experimental methods are combined to promote the identification of improved enzyme inhibitors. From a computational standpoint, the produced experimental data represent an invaluable resource to robustly validate the predictive power of the developed computational strategies and to apply them to identify novel promising molecules.

The MEDIATE initiative was clearly fostered by the pandemic crisis and was proposed as a shared strategy to respond to the urgent quest for effective therapies against SARS-CoV-2. We believe that this initiative can be a model of collaborative computing applied to virtual screening projects which can be easily applied to other targets (not necessarily focused on the antiviral field) to promote and accelerate hit identification and drug repurposing. It could be particularly fruitful in those fields (such as oncology, neurodegeneration, or cardiovascular diseases) on which countless scientific groups are performing computational studies and the possibility of synergistically combining their efforts might have a tremendous impact on the obtained results. In this regard, the ligands' libraries shared by MEDIATE can be considered as a valuable starting point (certainly expandable) to move forward standardized and thus shareable virtual screening projects.

The preliminary analyses here reported indicate that the submitted docking results are satisfactorily target-specific since the correlations between the score values of different targets are low and a vast majority of compounds are predicted to bind only one target. These outputs suggest that these docking results might be combined without generating biased consensus results as confirmed by the encouraging findings reported in Table 4. Nevertheless, it should be noted that almost all docking scores include a component which mostly depends on ligand's properties rather than on the specific ligand-target interactions. For example, almost all docking scores tend to improve with ligand's molecular size and the interactions elicited by polar ligands are often overestimated compared to hydrophobic ones. This suggests that the consensus analyses should be carefully designed to minimize as far as possible these biasing conditions by avoiding highly correlated docking scores and not truly target-specific scoring functions.

7. Expert opinion

As mentioned above, the pandemic crisis incredibly fostered the open science paradigm and various collaborative open projects were developed during the last few years. In fact, these joint projects are part of a much longer practice of shared initiatives involving the prediction of protein structures (CASP, the first edition dated back to 1994) [41], ligand-receptor interactions (D3R) 19] and physicochemical descriptors (SAMPL) [42]. Nevertheless, there is a significant difference between these projects and MEDIATE since they are all organized as challenges in which the researchers compete to provide the best performing results in terms of predictive ability. Clearly, this is a very productive strategy which exploits the natural competitive spirit and promoted remarkable enhancements in all the fields where it was applied. Nevertheless, the researchers are free to organize their researches in these challenge initiatives and this freedom, which underlies the expected competition among the involved research teams, leads to some dispersion of forces since the submitted simulations are not fully homogeneous and their synergistic exploitation is not really optimized. Such a strategy is particularly fruitful, but can become not completely suitable when the speed is a key factor in determining the success rate of a project.

In emergency situations, standardized conditions in which all researchers synergistically participate sharing the same input files in terms of protein structures and ligands' libraries are the best way to maximize the success rate of these initiatives. In this context, MEDIATE proposes a common platform that shares prepared and annotated input files and can become an example for similar initiatives of drug repurposing and virtual screening for other medicinal applications (not necessarily in urgent conditions). In such a more general context, the choice of the shared protein structures on which the simulations should be focused can play a key role. In the antiviral field to which MEDIATE is dedicated, the choice of the involved proteins is relatively simple because the number

of relevant targets (especially when focusing on the viral proteins) is limited and reasonably well-defined. The situation can become hugely more problematic when facing complex human diseases since in these cases the number of involved targets can incredibly grow, and their choice can be not univocal with subjective aspects.

Another interesting point concerning the updating of the shared protein targets. Indeed, especially when dealing with pandemic crises, one may imagine that the sharable structures will initially be mostly theoretical models which should be progressively replaced by experimentally resolved structures. While considering that resolved structures should conceivably provide better docking results, this protein replacement should not generate inhomogeneous results thus allowing a constant updating of the simulated proteins to share the best available structures.

Notably, in the future such shared collaborative simulations could benefit from the reasonable advancements in docking approaches concerning both the search engines and (especially) the scoring functions which markedly affect the reliability of the generated results. More importantly, these shared initiatives could benefit from the progress in artificial intelligence algorithms to enable an optimized combination of the submitted results so to develop highly predictive consensus predictive models.

With a view to fostering the worldwide participation, the MEDIATE initiative can be yet associated to a scientific challenge. Starting from standardized and shared starting data does not prevent to exploit the submitted simulations for a challenge among the participating groups, since the groups providing the best predictions could be awarded.

ORCID

Gianluca Palermo http://orcid.org/0000-0001-7955-8012

References

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Suran S, Pattanaik V, Draheim D. Draheim D frameworks for collective intelligence: a systematic literature review. ACM Comput Surv. 2020;53(1):1–36. Article 14. doi: 10.1145/3368986
2. Won JH, Lee H. Can the COVID-19 pandemic disrupt the current drug development practices? Int J Mol Sci. 2021;22(11):5457. doi: 10.3390/ijms22115457
3. Hollaway MJ, Dean G, Blair GS, et al. Tackling the challenges of 21st-century open science and beyond: a data science lab approach. Patterns. 2020;N Y(7):100103. doi: 10.1016/j.patter.2020.100103
4. Hufsky F, Lamkiewicz K, Almeida A, et al. Computational strategies to combat COVID-19: useful tools to accelerate SARS-CoV-2 and coronavirus research. Brief Bioinform. 2021;22(2):642–663. doi: 10.1093/bib/bbaa232
   •• This review describes the bioinformatics resources focused on the COVID-19 data
5. Aronskyy I, Masoudi-Sobhanzadeh Y, Cappuccio A, et al. Advances in the computational landscape for repurposed drugs against COVID-19. Drug Discov Today. 2021;26(12):2800–2815. doi: 10.1016/j.drudis.2021.07.026
   •• This review summarizes the approaches and obtained outcomes in repurposing campaigns for COVID-19
6. Ahsan MA, Liu Y, Feng C, et al. Bioinformatics resources facilitate understanding and harnessing clinical research of SARS-CoV-2. Brief Bioinform. 2021;22(2):714–725. doi: 10.1093/bib/bbaa416
7. Bernasconi A, Canakoglu A, Masseroli M, et al. A review on viral data sources and search systems for perspective mitigation of COVID-19. Brief Bioinform. 2021;22(2):664–675. doi: 10.1093/bib/bbaa359
8. Tanwar AS, Evangelatos N, Venne J, et al. Global open health data cooperatives cloud in an era of COVID-19 and planetary health. OMICS. 2021;25(3):169–175. doi: 10.1089/omi.2020.0134
9. Bittremieux W, Adams C, Laukens K, et al. Open science resources for the mass spectrometry-based analysis of SARS-CoV-2. J Proteome Res. 2021;20(3):1464–1475. doi: 10.1021/acs.jproteome.0c00929
10. Mei LC, Jin Y, Wang Z, et al. Web resources facilitate drug discovery in treatment of COVID-19. Drug Discov Today. 2021;26(10):2358–2366. doi: 10.1016/j.drudis.2021.04.018
11. xxxx. https://ec.europa.eu/info/research-and-innovation/research-area/health-research-and-innovation/coronavirus-research-and-innovation/covid-research-manifesto_en
12. Lagarde N, Goldwaser E, Pencheva T, et al. A free web-based protocol to assist structure-based virtual screening experiments. Int J Mol Sci. 2019;20(18):4648. doi: 10.3390/ijms20184648
13. Murail S, de Vries SJ, Rey J, et al. SeamDock: an interactive and collaborative online docking resource to assist small compound molecular docking. Front Mol Biosci. 2021;8:716466. doi: 10.3389/fmolb.2021.716466

14. Tse EG, Klug DM, Todd MH. Open science approaches to COVID-19 [version 1; peer review: 2 approved]. F1000Res. 2020;9:1043.
   • This review describes the open/collaborative project focused on COVID-19
15. The COVID Moonshot Consortium, Chodera J, Lee A, et al. Open science discovery of oral non-covalent sarS-CoV-2 main protease inhibitors. ChemRxiv. 2021.
   • A relevant example of collaborative project for COVID-19
16. JEDI COVID-19 Grand Challenge. https://www.jedi.foundation/covid19challenge
17. MMV COVID Box: https://www.mmv.org/mmv-open/archived-projects/covid-box
18. Zimmerman MI, et al. Citizen Scientists Create an Exascale Computer to Combat COVID-19. bioRxiv 2020.06.27.175430; 2020. doi: 10.1101/2020.06.27.175430
19. Parks CD, Gaieb Z, Chiu M, et al. D3R grand challenge 4: blind prediction of protein–ligand poses, affinity rankings, and relative binding free energies. J Comput Aided Mol Des. 2020;34(2):99–119. doi: 10.1007/s10822-020-00289-y
   • A well-known example of challenge in the field of docking simulations
20. Grosjean H, Işık M, Aimon A, et al. SAMPL7 protein-ligand challenge: a community-wide evaluation of computational methods against fragment screening and pose-prediction. J Comput Aided Mol Des. 2022;36(4):291–311. doi: 10.1007/s10822-022-00452-7
21. Yin J, Henriksen NM, Slochower DR, et al. Overview of the SAMPL5 host–guest challenge: are we doing better? J Comput Aided Mol Des. 2017;31(1):1–19. doi: 10.1007/s10822-016-9974-4
22. Ghose AK, Crippen GM. Atomic physicochemical parameters for three-dimensional-structure-directed quantitative structure-activity relationships. 2. Modeling dispersive and hydrophobic interactions. J Chem Inf Comput Sci. 1987;27(1):21–35. doi: 10.1021/ci00053a005
23. Ertl P, Rohde B, Selzer P. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. J Med Chem. 2000;43(20):3714–3717. doi: 10.1021/jm000942e
24. Harder E, Damm W, Maple J, et al. OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. J Chem Theory Comput. 2016;12(1):281–296. doi: 10.1021/acs.jctc.5b00864
25. van Vlijmen H, Ortholand JY, Li VM-J, et al. The European Lead Factory: an updated HTS compound library for innovative drug discovery. Drug Discov Today. 2021;26(10):2406–2413. doi: 10.1016/j.drudis.2021.04.019
26. Gervasoni S, Vistoli G, Talarico C, et al. A comprehensive mapping of the druggable cavities within the SARS-CoV-2 therapeutically relevant proteins by combining pocket and docking searches as implemented in pockets 2.0. Int J Mol Sci. 2020;21(14):5152. doi: 10.3390/ijms21145152
27. Corona A, Wycisk K, Talarico C, et al. Natural compounds inhibit SARS-CoV-2 nsp13 unwinding and ATPase enzyme activities. ACS Pharmacol Transl Sci. 2022;5(4):226–239. doi: 10.1021/acsptsci.1c00253
28. Fraser BJ, Beldar S, Seitova A, et al. Structure and activity of human TMPRSS2 protease implicated in SARS-CoV-2 activation. Nat Chem Biol. 2022;18(9):963–971. doi: 10.1038/s41589-022-01059-7
29. Le Guilloux V, Schmidtke P, Tuffery P. Fpocket: an open source platform for ligand pocket detection. BMC Bioinf. 2009;10(1):168. doi: 10.1186/1471-2105-10-168
30. Pedretti A, Mazzolari A, Gervasoni S, et al. The VEGA suite of programs: an versatile platform for cheminformatics and drug design projects. Bioinformatics. 2021;37(8):1174–1175. doi: 10.1093/bioinformatics/btaa774
31. Torrens-Fontanals M, Peralta-García A, Talarico C, et al. SCoV2-MD: a database for the dynamics of the SARS-CoV-2 proteome and variant impact predictions. Nucleic Acids Res. 2022;50(D1):D858–D866. doi: 10.1093/nar/gkab977
32. Chu H, Chan JF, Yuen TT, et al. Comparative tropism, replication kinetics, and cell damage profiling of SARS-CoV-2 and SARS-CoV with implications for clinical manifestations, transmissibility, and laboratory studies of COVID-19: an observational study. Lancet Microbe. 2020;1(1):e14–e23. doi: 10.1016/S2666-5247(20)30004-5
33. Zhang L, Lin D, Sun X, et al. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. Science. 2020;368(6489):409–412. doi: 10.1126/science.abb3405
34. Shin D, Mukherjee R, Grewe D, et al. Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. Nature. 2020;587(7835):657–662. doi: 10.1038/s41586-020-2601-5
35. Acharya A, Agarwal R, Baker MB, et al. Supercomputer- based ensemble docking drug discovery pipeline with application to Covid-19. J Chem Inf Model. 2020;60(12):5832–5852. doi: 10.1021/acs.jcim.0c01010
36. Mansouri K, Karmaus AL, Fitzpatrick J, et al. CATMoS: collaborative acute toxicity modeling suite. Environ Health Perspect. 2021;129(4):47013. doi: 10.1289/EHP8495
37. Mansouri K, Kleinstreuer N, Abdelaziz AM, et al. CoMPARA: collaborative modeling project for androgen receptor activity. Environ Health Perspect. 2020;128(2):27002. doi: 10.1289/EHP5580
38. Beccari AR, Vistoli G. Exscalate4cov: innovative High Performing Computing (HPC) strategies to tackle pandemic crisis. Int J Mol Sci. 2022;23(19):11576. doi: 10.3390/ijms231911576
39. Kuzikov M, Constanzi E, Reinshagen J, et al. Identification of Inhibitors of SARS-CoV-2 3CL-Pro enzymatic activity using a small molecule in vitro repurposing screen. ACS Pharmacol Transl Sci. 2021;4(3):1096–1110. doi: 10.1021/acsptsci.0c00216
40. Zaliani A, Vangeel L, Reinshagen J, et al. Cytopathic SARS-CoV-2 screening on VERO-E6 cells in a large-scale repurposing effort. Sci Data. 2022 Jul 13;9(1):405. x
41. Protein Structure Prediction Center: https://predictioncenter.org/
42. Rodriguez SA, Tran J, Sabatino SJ, et al. Predicting octanol/water partition coefficients and pKa for the SAMPL7 challenge using the SM12, SM8 and SMD solvation models. J Comput Aided Mol Des. 2022;36(9):687–705. doi: 10.1007/s10822-022-00474-1