Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Targeting the SARS-CoV-2 spike glycoprotein prefusion conformation: virtual screening and molecular dynamics simulations applied to the identification of potential fusion inhibitors

Alice Romeo¹, Federico Iacovelli¹, Mattia Falconi*²

1. Introduction

The coronavirus disease 2019 (COVID-19) pandemic has rapidly become an international public health threat, which is causing an inconceivable loss of lives and a severe socio-economic instability worldwide. The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), identified as the causal agent of this disease, is an enveloped single-stranded positive-sense RNA virus assigned to the species Severe acute respiratory syndrome-related coronavirus, genus Betacoronavirus, family Coronaviridae, and was firstly identified from a cluster of pneumonia cases in Wuhan (China) on December 2019 (Zhou et al., 2020). On January 30, 2020 the World Health Organization (WHO) declared the coronavirus epidemic as a Public Health Emergency of International Concern. WHO raised the threat of the coronavirus epidemic to the world level to “very high” on February 28, and on March 11 the COVID-19 was officially declared a worldwide pandemic. On March 13, WHO stated that Europe is becoming the new epicentre of the pandemic. The main cause of the uncontrollable spread of this disease is the lack of specific antiviral medications, though several groups in the world are making a lot of effort in trying to speed up this research. Although the production of vaccines could be the most powerful weapon against SARS-CoV-2 spread and invasion, it will take quite a long time to develop and clinically test their safety. Thus, as an emergency measure, it would be fundamental to identify or develop antiviral molecules not only to treat the COVID-19 but also to prevent its further transmission.

Although no antiviral treatment has yet been approved, several therapeutic approaches have been proposed, based on the administration of molecules developed for treating other diseases as the combination of lopinavir/ritonavir, chloroquine and hydroxychloroquine (Kupferschmidt and Cohen, 2020). Preclinical studies also suggested that remdesivir, a RNA polymerase inhibitor developed against Ebola, could be effective for both prophylaxis and therapy of human coronavirus infections, as suggested by results obtained testing its activity in a rhesus macaque model of MERS-CoV infection (de Wit et al., 2020). Finally, an investigation led by the Istituto Nazionale Tumori, Fondazione Pascale di Napoli is focused on the use of tocilizumab, a humanized IgG1 monoclonal antibody, directed against the IL-6 receptor and commonly used in the treatment of rheumatoid arthritis, to prevent the cytokine storm syndrome caused by the SARS-CoV-2 infection.

* Corresponding author at: University of Rome “Tor Vergata”, Department of Biology, Via della Ricerca Scientifica 1, 00133, Rome, Italy.
E-mail address: falconi@uniroma2.it (M. Falconi).
¹These authors contributed equally to this work and should be considered co-first authors.

https://doi.org/10.1016/j.virusres.2020.198068
Received 10 April 2020; Received in revised form 3 June 2020; Accepted 16 June 2020
Available online 18 June 2020
0168-1702/ © 2020 Elsevier B.V. All rights reserved.
purposing with the aim to subsequently virus entry into host cells. The S protein is a trimeric class I fusion protein that exists in a metastable prefusion conformation, which undergoes substantial structural rearrangements to fuse the viral membrane with the host cell membrane (Bosch et al., 2003; Li, 2016). The S protein further subdivides into two distinct functional subunits: the S1 subunit is involved in host cell receptor recognition, while the S2 mediates viral fusion to host cells. Fusion process is triggered when the S1 subunit binds to a specific cell receptor. Receptor binding destabilizes the pre-fusion trimer, resulting in shedding of the S1 subunit and transition of the S2 subunit to a stable postfusion conformation. The latter is characterized by the assembly of its heptad repeats 1 (HR1) and 2 (HR2) regions into a 6-helix bundle structure that allows viral fusion and entry into host cells (Bosch et al., 2003; Walls et al., 2017). The complete three-dimensional structure of the human SARS-CoV-2 spike in postfusion conformation has not yet been published but recently Wrapp and coworkers (Wrapp et al., 2020) determined the first 3.5 Å resolution cryo-electron microscopy (cryo-EM) structure of the SARS-CoV-2 S trimer in the pre-fusion conformation. The study provides biophysical and structural evidences that the SARS-CoV-2 S protein binds the ACE2 receptor with higher affinity than does the closely related severe acute respiratory syndrome coronavirus (SARS-CoV) S protein. Additionally, the authors tested several published SARS-CoV RBD-specific monoclonal antibodies and found no appreciable binding to the SARS-CoV-2 S protein, suggesting that antibody cross-reactivity may be limited between the RBDs of these two viruses (Wrapp et al., 2020).

Because of the crucial function of the S protein in the SARS-CoV-2 infection process, this structural component could represent a target for antibody or small-molecule mediated neutralization, and characterization of the pre-fusion S structure allowed to get atomic-level information essential to guide the design and development of novel therapeutic molecules.

Previous data obtained for the human respiratory syncytial virus (hRSV) F protein, a class I viral fusion glycoprotein sharing structural similarities with most of the lower respiratory tract viruses’ fusion glycoproteins, indicated that newly generated escape mutations for evaluated small-molecule fusion inhibitors were clustered inside the central cavity of the F protein in pre-fusion conformation (Battles and McLellan, 2019; Lopes et al., 2020; Yan et al., 2014). Crystallographic studies later confirmed that these small-molecules bind within the central cavity, contacting from two to three lobes of the trimeric binding pocket and interacting with residues of the fusion peptide and the second heptad repeat (Battles et al., 2016). Functional studies involving soluble F proteins (Samuel et al., 2015) and membrane-bound F proteins (Battles et al., 2016), further demonstrated that the evaluated small-molecules act as antagonists that prevent the protein transition to the postfusion conformation. To date, all known small-molecule fusion inhibitors bind in the same pocket and have the same mechanism of action (DeVincenzo et al., 2014; Stevens et al., 2018). These data suggest that, also in the case of SARS-CoV-2, targeting the central cavity of the spike glycoprotein in pre-fusion conformation could be a valuable approach for hindering its pre-fusion to postfusion transition and consequently virus entry into host cells.

In this paper, we have used the SARS-CoV-2 S trimer structure in pre-fusion conformation (Wrapp et al., 2020) to perform a drug repurposing with the aim to find and suggest small-molecule fusion inhibitors for clinical use. Repurposing of drugs already approved, or still in clinical trial, represents an accelerated route for drug discovery, because existing drugs have already established clinical and pharmacokinetic data. We have carried out a virtual screening (VS) simulation using 8,770 FDA drugs extracted from the DrugBank database (https://www.drugbank.ca/) (Wishart et al., 2018). The structure of the spike glycoprotein (Wrapp et al., 2020) has been completed using modeling techniques and used for the identification of available compounds strongly interacting inside its inner cavity (Fig. 1). Furthermore, the stability of selected protein-drug complexes has been tested using classical molecular dynamics (MD) simulations and the free energy of interaction of these molecules has been finely evaluated through the molecular mechanics energies combined with generalized Born and surface area continuum solvation (MM/GBSA) method and per-residue energy decomposition analysis (Gehedeen and Ryde, 2015). Finally, alignments between the sequence of the spike glycoprotein used in our study and sequences of all betacoronavirus spike glycoproteins reported in the GenBank database (Clark et al., 2016), including recently deposited sequences of different SARS-CoV-2 isolates, have been carried out to evaluate the presence of variability within the protein region analyzed by this simulation research.

As a result, considering the available scientific information, a selected set of drugs has been suggested as SARS-CoV-2 fusion inhibitors, with the hypothesis that the use of these compounds can be extended also to other related viruses.
2. Methods

2.1. Molecular modeling

Cryo-EM structure of the SARS-CoV-2 spike glycoprotein in prefusion conformation has been obtained from the PDB database (PDB ID: 6vsb) (Wrapp et al., 2020). The trimer structure, composed by three identical monomers (named A, B, C), has been completed modeling non-terminal missing loops using the SWISS-MODEL web-server (Waterhouse et al., 2018). The modeling procedure mainly involved regions of the S1 subunit of the glycoprotein (residues: 67–78, 96–98, 143–155, 177–186, 247–260, 329–334, 444–448, 455–490, 501–502, 621–639, 673–686; of monomer A, residues: 67–80, 142–154, 177–186, 210–216, 243–262, 444–448, 455–459, 474–486, 501–502, 621–637, 673–686; of monomer B and residues: 67–79, 96–98, 141–156, 177–186, 246–260, 444–448, 455–459, 472–486, 499–502, 621–640, 673–686; of monomer C), and also two regions of the S2 subunit (residues 812–814 and 829–852 of monomers A, B and C).

Although the modeled regions were not close to the central cavity of the spike glycoprotein, they have been rebuilt since it is not possible to accurately simulate and analyze a protein with missing parts using classical MD simulation. In fact, during the simulation, the protein would be subjected to unnatural fluctuations, misleading the subsequent analyses.

2.2. Virtual screening procedure

A drug library, containing 8,770 compounds in MOL2 format, has been extracted from the DrugBank database v. 5.1.5 (Wishart et al., 2018) (https://www.drugbank.ca/). Spike glycoprotein receptor and drug structure files have been converted into pdbqt format using the prepare_receptor4.py and the prepare_ligand4.py tools of the AutoDockTools4 software (Morris et al., 2009; Sanner, 1999). Seven drug structures have been excluded from the analysis, since prepare_ligand4 could not generate their PDBQT files, while other eight structures, containing Si atoms, have been excluded due to the absence of force-field parameters. The 8755 molecular docking simulations have been performed using an in-house parallelized version of the Autodock Vina 1.1.2 program (Trott and Olson, 2010). One docking simulation, including ten docking runs, has been carried out for each drug. For each docking simulation, the Autodock Vina program selects 10 binding poses representing the cluster centroids of all the different conformations, generated in each run using the Lamarckian Genetic Algorithm. A box of size x = 23.25 Å, y = 24.38 Å, z = 25.88 Å has been placed over the HR1 internal region (residues 897–920) of the spike glycoprotein A monomer. 15 receptor residues side-chains around the selected binding site have been considered as flexible (Y904, F906, N907, I909, V911, T912, Q913, N914, V915, E1092, Q1106, N1108 of the monomer A and R1091, E1092, F1121 of the monomer A). Binding energies have been calculated as an average of the best poses obtained from three repeated docking simulations.

The sequential molecular docking simulations of the 4 top drugs, obtained after re-evaluating the compounds ranking, have been performed for the first three docked compounds applying the same parameters already described for the 3 compounds molecular docking. Molecular docking of the fourth compound, instead, was performed using a box of size x = 27.00 Å, y = 30.38 Å, z = 25.50 Å centred between the HR1 internal regions of the monomers B and C, selecting 12 receptor side chains as flexible (I909, T912, E1092, Q1106, R1107, N1108, F1109 of the monomer B and Y904, R905, N907, Q1036, K1038 of the monomer C).

All molecular docking simulations took about 30’ and have been performed using the ENEA HPC cluster CRESCo6 (Ponti et al., 2014).

2.3. Molecular docking simulations of multiple compounds

Two sequential molecular docking simulations have been performed for phthalocyanine and hypericin, the two top-ranking drugs, in order to dock a second and a third molecule of the compounds inside the spike glycoprotein binding pocket. The best complexes obtained in the virtual screening first docking run have been used as receptors, converting the structures into pdbqt format using the prepare_receptor4.py tool of the AutoDockTools4 software (Morris et al., 2009; Sanner, 1999). The two molecular docking simulations, each including ten docking runs, have been carried out using the Autodock Vina 1.1.2 program (Trott and Olson, 2010). A box of size x = 25.88 Å, y = 24.00 Å, z = 25.88 Å has been centred over the HR1 internal region of the spike glycoprotein monomer B, selecting 15 residues side-chains around this binding site as flexible (Y904, F906, N907, I909, V911, T912, Q913, N914, V915, E1092, Q1106, N1108 of the monomer B and R1091, E1092, F1121 of the monomer A). Finally, a box of size x = 24.75 Å, y = 25.50 Å, z = 25.88 Å has been placed over the HR1 internal region belonging to the monomer C, selecting 15 side chains as flexible (Y904, F906, N907, I909, V911, T912, Q913, N914, V915, E1092, Q1106, N1108 of the monomer C and R1091, E1092, F1121 of the monomer B). Binding energies have been calculated as an average of the best poses obtained from three repeated docking simulations.

The sequential molecular docking simulations of the 4 top drugs, obtained after re-evaluating the compounds ranking, have been performed for the first three docked compounds applying the same parameters already described for the 3 compounds molecular docking. Molecular docking of the fourth compound, instead, was performed using a box of size x = 27.00 Å, y = 30.38 Å, z = 25.50 Å centred between the HR1 internal regions of the monomers B and C, selecting 12 receptor side chains as flexible (I909, T912, E1092, Q1106, R1107, N1108, F1109 of the monomer B and Y904, R905, N907, Q1036, K1038 of the monomer C).

The two complexes obtained with multiple docking of phthalocyanine and hypericin drugs have been simulated using classical molecular dynamics. Topologies and coordinates files of the input structures have been generated using the tLeap module of the AmberTools 19 package (Case et al., 2018). The spike glycoprotein has been parametrized using the ff19SB force field (Tian et al., 2020), while parameters for the two top-ranking drugs have been generated using the antechamber module of the AmberTools 19 package (Case et al., 2018) and the general Amber force field (Wang et al., 2004). Each spike glycoprotein, complexed with three drugs, has been inserted into a rectangular box of TIP3P water molecules (Jorgensen et al., 1983), setting a minimum distance of 12.0 Å from the box sides and neutralizing the solution with 0.15 mol/L of NaCl ions. In order to remove unfavourable interactions, structures have been subjected to four minimization cycles, each composed by 500 steps of steepest descent minimization followed by 1500 steps of conjugated gradient minimization. A starting restraint of 20.0 kcal·mol⁻¹·Å⁻² has been imposed on protein and ligand atoms and subsequently reduced and removed in the last minimization cycle. Systems were gradually heated from 0 to 300 K in a NVT ensemble over a period of 2.0 ns using the Langevin thermostat (Goga et al., 2012), imposing a starting restraint of 0.5 kcal·mol⁻¹·Å⁻² on each protein and ligand atom, which was decreased every 500 ps in order to slowly relax the system. The systems were simulated in an isobaric-isothermal (NPT) ensemble for 2.0 ns, imposing a pressure of 1.0 atm using the Langevin barostat (Aoki et al., 2004) and fixing the temperature at 300 K. Covalent bonds involving hydrogen atoms have been constrained using the SHAKE algorithm (Ryckaert et al., 1977). A production run of 30 ns has been performed for both systems with a timestep of 2.0 fs, using the pmemd.cuda module of the AMBER 18 software (Case et al., 2018). The PME method was used for calculating long-range interactions while a cut-off of 9.0 Å has been set for short-range interactions (Darden et al., 1993). System coordinates have been saved every 1000 steps.

2.5. Trajectory analysis

RMSD analysis for the spike glycoprotein has been performed using the rms module of the GROMACS 2019 analysis tool (Abraham et al., 2015). Hydrogen bond analyses have been carried out using the HBonds plugin of the software VMD 1.9.3 (Humphrey et al., 1996). Generalized
Born and surface area continuum solvation (MM/GBSA) analysis (Genheden and Ryde, 2015) and per-residue energy decomposition analysis have been performed over the last 10 ns of both simulations, using the MMPBSA.py.MPI program implemented in the AMBER18 software (Case et al., 2018) on 3 nodes of the ENEA HPC cluster CRESC06 (Ponti et al., 2014). Decomposition analysis has been performed on residues surrounding each drug, located within a cut-off distance of 12 Å.

3. Results and discussion

3.1. Virtual screening

A Virtual Screening (VS) analysis (drug repurposing) of a set of 8755 structures of approved, experimental and investigational drugs, extracted from the DrugBank database (https://www.drugbank.ca/) (Wishart et al., 2018), has been performed in order to find compounds that, entering the inner cavity of the protein and blocking the HR1 conformational changes, could act as spike glycoprotein fusion inhibitors.

Using an in-house parallelized version of the AutoDock Vina software (Trott and Olson, 2010), we evaluated the binding affinity of each drug for one of the three HR1 internal regions of the spike glycoprotein (residues 897–920). The ten top-ranking compounds obtained from the VS analysis are reported in Fig. 2 and in Table S1. Interestingly, three of these compounds have already been reported as possible antiviral agents, targeting different types of enveloped viruses. In particular, the compounds 31h-phthalocyanine (from here simply referred to as phthalocyanine) and hypericin represent the two most appealing molecules. In fact, they show a significantly high affinity for the HR1 region and, additionally, hypericin and different phthalocyanine derivatives have been already shown to prevent human immunodeficiency virus (HIV) infections, interfering with the surface glycoprotein gp120 binding and fusion to the host cell (François et al., 2009; Lenard et al., 1993; Vzorov et al., 2003). Furthermore, these two compounds, or their derivatives, also possess different antiviral activities towards other viral species, as it will be discussed below in Section 3.5, and hypericin has also been recently shown to inhibit the infectious bronchitis virus (IBV), an avian coronavirus, acting through a mechanism involving inhibition of the virus-induced cellular apoptosis (Chen et al., 2019b).

3.2. Multiple molecular docking simulations within the spike glycoprotein inner cavity

Considering the trimeric structure of the spike glycoprotein, we propose an inhibitory mechanism by which at least three drug molecules could enter and arrange inside the inner cavity, binding the three HR1 internal regions (Wrapp et al., 2020). The concurrent binding of three molecules could generate a strong steric hindrance inside the pocket, preventing the conformational changes of the HR1 regions and hence the formation of the 6-helix bundle stabilizing the postfusion conformation (Walls et al., 2017).

In order to evaluate our hypothesis, starting from the top ranking complexes obtained from the VS procedure, we evaluated the binding affinity of two phthalocyanine or two hypericin molecules for the other two HR1 internal regions of the spike glycoprotein. The average binding energies obtained for each molecule have been reported in Table 1. It can be observed that even if the binding affinities of the second and third bound molecules are lower than that obtained for the first, due to the reduction in the conformational space inside the pocket, they still remain particularly suitable for establishing valuable interactions. Both compounds contact a high number of residues within the pocket, both from the HR1 internal regions and from upward and downward regions of the spike, creating a connection between the HR1s and surrounding regions of the inner cavity (Fig. 3).

Phthalocyanine molecules arrange in a symmetric orientation (Fig. 3 A–B), facing the inner cavity of the pocket and establishing 11 hydrophobic interactions with surrounding residues. In particular, stacking interactions are settled between the isoindole groups of phthalocyanines and aromatic side-chains of the protein (Fig. 3 B). The three phthalocyanines can also set up hydrophobic interactions within themselves, increasing the stability of the multiple complex (Fig. 3 B–C).

Hypericin molecules, on the other hand, bind to the HR1 regions in an alternate orientation (Fig. 3 D–E), establishing 7 hydrophobic interactions and 4 hydrogen bonds with the surrounding residues. The hypericin molecules also set up contacts within themselves, interacting through hydrogen bonds established between their hydroxyl groups (Fig. 3 F). However, the higher affinity showed by phthalocyanine for the HR1 regions suggests that hydrophobic interactions can drive the formation of extremely favourable associations inside the pocket, hypothesis supported by the presence of many aromatic moieties in all the compounds showing higher affinity for this internal cavity (Fig. 2).

It is difficult to hypothesize what the location of these compounds might be in an intermediate or postfusion form of the spike, since complete postfusion structures are still not available. From a recently published incomplete structure of the spike postfusion core (PDB ID: 6lxt) (Xia et al., 2020) we found that only three of the key cavity residues, involved in interactions with the compounds, are present in the spike 6-helix bundle structure (residues 912–914). Nevertheless, since our aim was to prevent virus entry in an earlier phase, directly blocking the prefusion conformation, potential interactions of the drugs with structures in conformational progress were not taken into further consideration since they would not add details to improve our model.

3.3. MD simulations and MM/GBSA interaction energies evaluation

To further assess the affinity of the top two drugs for the spike pocket we performed 30 ns MD simulations of both multiple docking complexes to accurately evaluate, through the MM/GBSA method (Genheden and Ryde, 2015), the interaction energies between the docked compounds and the receptor. MM/GBSA analysis confirmed the high affinity of both drugs for the binding pocket (Table 2). In particular, MM/GBSA results highlighted that Van der Waals, electrostatic and non-polar solvation energy terms favourably contribute to the binding energy. The Van der Waals term, accounting the strength of nonpolar interactions derived from the binding of the drugs to the pocket, gives the highest contribution to the binding energy, showing extremely negative values. On the other hand, highly positive polar solvation energy terms have been obtained in all the calculations. This term, representing the energy associated with dissolving these drugs in the solvent, depict their preference for a hydrophobic environment.

Highly favourable Van der Waals energies have been obtained for the phthalocyanine drugs, while electrostatic interactions with surrounding residues play only a marginal role indicating that the massive hydrophobic contacts established by these molecules within the pocket are crucial for the stability of the complex. On the other hand, electrostatic interactions are important for the stability of the hypericin drugs, although the hydrophobic interactions still play a significant role for their binding affinity.

Hydrogen bond analysis, performed for both the simulations, further supports these results, showing that the presence of the three hypericin drugs leads to an increase in the hydrogen bonds network established within the pocket. In fact, these molecules can stably set up from 1 to 4 hydrogen bonds (Fig. S1, lower panel), mainly with residues P1090, R1091 and E1092 of the monomer A, N907, E1092 and Q1113 of the monomer B and P1090 of the monomer C, as well as with themselves. Conversely, polar interactions play only a minor role in phthalocyanine drugs binding to the pocket, since none of the three molecules established valuable hydrogen bonds during the simulation time (Fig. S1, upper panel).

Per-residue energy decomposition analysis (Genheden and Ryde,
was also carried out for each drug to highlight the independent contribution of specific pocket residues to the binding, and to dynamically describe the compounds interactions within the pocket (Table S2). The analysis showed that the drugs can establish from 9 to 15 medium- to high-affinity contacts with residues belonging to the internal HR1 regions (residues 897–920) and with other surrounding residues inside the cavity. This evidence further supports the hypothesis that the drugs may act like anchors between surrounding internal regions. Moreover, the results also confirmed that the interactions established by the three phthalocyanines are mostly hydrophobic, while the three hypericin molecules can also set up valuable electrostatic interactions with most of the contacted residues, although the hydrophobic contribution is still much stronger.

The MD and MM/GBSA results mainly confirmed the interaction patterns described by the molecular docking simulations, indicating that highly stable binding poses have been obtained. These molecules improve the number of hydrophobic interactions inside the pocket, creating a non-polar network and the establishment of strong steric hindrance within the cavity that may likely hamper the subsequent conformational changes leading to the formation of the postfusion conformation of the spike. Interestingly, during the simulations it has been also observed that the first and second docked phthalocyanine or hypericin molecules maintain their starting binding poses interacting with the HR1 internal regions of the monomer A and B, while the third added phthalocyanine or hypericin compound moves towards the centre of the pocket, simultaneously contacting all the HR1 internal

Table 1

| Compound    | Binding energy #1 (kcal/mol) | Binding energy #2 (kcal/mol) | Binding energy #3 (kcal/mol) |
|-------------|-------------------------------|-------------------------------|-------------------------------|
| phthalocyanine | -16.3                         | -16.0                         | -14.6                         |
| hypericin    | -15.1                         | -14.8                         | -12.2                         |

Fig. 2. Best 10 high affinity compounds identified through the Virtual Screening procedure. 2D representations have been obtained from the DrugBank database (Wishart et al., 2018) (https://www.drugbank.ca/). Drugs order follows that of Table S1.
The arrangement of these compounds is expected to create a hard impairment to the motion of the HR1 regions, further contributing to the stabilization of the spike prefusion conformation. In conclusion, all these results sustain the hypothesis that the presence of these drugs inside the pocket may block the HR1 regions, preventing the huge spike conformational changes that permit the SARS-CoV-2 entry into the target cells.

### 3.4. Virtual screening ranking reweighting

The ranking of the compounds obtained through the VS procedure (Table S1) has been reweighted based on the literature retrieved for each compound (Table 3, and Fig. S2 and S3). In particular, we conceived a scoring method which allowed us to evaluate a reformulated ranking based on: a) the obtained binding energies, b) known side effects and physiological effects of compounds and c) previously identified antiviral properties.

Briefly, side effects have been classified using a score ranging between -5 and +5 (named S(SE)). In particular, compounds interfering with other drugs have been penalized, compounds with mild side effects have been assigned a positive score and compounds with unknown effects received a score of 0. Physiological effects of the compounds also received a score ranging between -5 and 5 (named S(PE)), according to which compounds with unfavourable physiological activities were negatively weighted, compounds with predicted or evaluated antiviral or anti-inflammatory activity received a value between +1 and +5 and compounds with unknown or neutral physiological activities obtained a value of 0.

Drugs previously evaluated in clinical trials have been positively weighted, computing the clinical trial score (named S(Trials)) as follows:

\[
S(\text{Trials}) = 5 + (1 \times n_{\text{AT}}) + (0.1 \times n_{\text{PE}}) + (0.1 \times n_{\text{PhaseI}}) + (0.2 \times n_{\text{PhaseII}}) + (0.3 \times n_{\text{PhaseIV}})
\]

Where \(n_{\text{AT}}\) indicates the number of clinical trials where the compound has been evaluated for its antiviral action, \(n_{\text{PE}}\) the number of trials with other clinical targets and \(n_{\text{PhaseI}}, n_{\text{PhaseII}}, n_{\text{PhaseIV}}\), the number of trial in each experimental phase, as reported on the U.S. National Library of Medicine (NLM) Clinical Trials database (https://clinicaltrials.gov). A
maximum value of 0.5, 1.0 and 1.5 was allowed for the last three terms of the equation to balance the ranking, and a maximum value of 10 for the overall $S$(Trials) score. Compounds with antiviral properties, even if directed towards viruses and/or targets different from the one chosen in our research, have been highly favoured in the final score. This property can add an extra value to the drugs, assuming that they could affect other SARS-CoV-2 targets in addition to the spike cavity. Furthermore, in particular in the case of phthalocyanine, the antiviral activity was experimentally observed but the specific viral targets were not completely understood (Vzorov et al., 2003), so the drug interaction within the inner cavity could provide a reason for this still undefined antiviral activity. For the $S$(Antiviral) calculation, compounds with known antiviral activity have been assigned a score of +10, while compounds for which this activity was only hypothesized, but not experimentally assessed, were assigned a score of +5. If the compound showed activity on enveloped viruses, a +5 was added to the score. Finally, according to confirmed data, a score between +10 and +20 was given to those compounds for which the reported inhibitory mechanism involved the binding of spike glycoproteins and/or the inhibition of virus fusion to host cells. For each compound, the average binding energy (named $S$(Binding)) has been calculated from three repeated molecular docking simulations and summed to the obtained score. Finally, a +2 wt was added to the best 5 compounds obtained from the VS procedure (named $S$(VS)), as reported in Table S1. Concluding, the final score for each compound was computed using the following equation:

$$S_{\text{final}} = S(\text{SE}) + S(\text{PE}) + S(\text{Trials}) + S(\text{Antiviral}) + S(\text{Binding}) + S(\text{VS})$$

The reweighted ranking reported in Table 3 shows that four top ranking compounds (i.e. phthalocyanine, hypericin, TMC-647055 and quarfloxin), evaluated combining of all the analyzed data, represent the most promising drugs and are suggested as a possible treatment against SARS-CoV-2 infection.

### Table 3

| Name            | S(SE) | S(PE) | S(Trials) | S(Antiviral) | S(Binding) | S(VS) | S(final) |
|-----------------|-------|-------|-----------|--------------|------------|-------|----------|
| phthalocyanine  | 4     | 0     | 5.3       | 35           | 16.3       | 2     | 62.6     |
| hypericin       | 2     | 3     | 8.2       | 25           | 15.1       | 2     | 55.3     |
| TMC-647055      | 5     | 5     | 10.0      | 15           | 12.5       | 2     | 49.5     |
| quarfloxin      | 5     | 5     | 5.6       | 5            | 12.6       | 2     | 35.2     |
| tepoindib       | 5     | 0     | 7.1       | 0            | 12.0       | 0     | 24.1     |
| laniquidar      | 5     | 0     | 5.2       | 0            | 12.8       | 0     | 23.0     |
| tadalaflf       | 2     | -3    | 10.0      | 0            | 12.4       | 0     | 21.4     |
| ergotamine      | -1    | 0     | 5.9       | 0            | 13.2       | 2     | 20.1     |
| NNI-10311795    | 1     | 5     | 0.0       | 0            | 13.1       | 0     | 19.1     |
| TZZP16          | 0     | 0     | 0.0       | 0            | 12.7       | 0     | 12.7     |

3.5. Known or hypothesized antiviral properties of the suggested compounds

To highlight the clinical potential of the four selected compounds, we provide a brief description for each drug, strictly focused on known or hypothesized antiviral properties. The listed effects, although not directed towards SARS-CoV-2 and not representing an evidence of interaction with this virus, may contribute to the general antiviral properties of the drug cocktail.

- Phthalocyanines are dye molecules commonly used for optical, electrical and catalytic applications (Sorokin, 2013; Wright, 2001). Different phthalocyanine derivatives have also been evaluated for their efficacy as virucidal agents and as fusion inhibitors (François et al., 2009; Vzorov et al., 2003, 2007). Given their photochemical properties, phthalocyanines also showed a phototoxic virucidal action on different enveloped viruses (Korneev et al., 2019; Nikolaeva-Glomb et al., 2017; Smetana et al., 1994). Concerning the safety of this molecule, no cell toxicity has been reported and an in vivo study on mice showed a good toxicity profile and no tissue damage (Styczynski et al., 2015).

- Hypericin is a natural naphthodiantrone compound mainly derived from plants of the genus *Hypericum*, particularly *Hypericum perforatum* (St. John’s Wort), and widely known for its antidepressant activity (Jenďelovská et al., 2016; Karioti and Bilia, 2010). Hypericin has also shown specific antiviral activity on a wide variety of enveloped viruses, mainly related on the viral assembly and release inhibition (Rubin et al., 2005). Hypericin was also reported to act as a fusion inhibitor for different viruses (Rubin et al., 2005; Lenard et al., 1993; Stevenson and Lenard, 1993). Hypericin and *H. perforatum* extracts have also been shown to inhibit the cellular apoptosis induced by the avian coronavirus IBV, and to attenuate the pathological injuries caused by this virus in the trachea and kidney of chickens (Chen et al., 2019a, b). Side effects reported for St. John’s Wort treatment include skin rash, dizziness, headache, tiredness and difficult sleep.

- TMC-647055 is a macrocyclic indole, non-nucleoside inhibitor of the hepatitis C virus (HCV) NS5B RNA-dependent RNA polymerase (Cummings et al., 2014; Devogelaere et al., 2012; Vendeville et al.,...
Clinical trials have highlighted its promising antiviral properties against HCV, as well as its good safety and tolerability profiles, with limited toxicity and off-target activities. In combination with the TMC435 protease inhibitor (also known as simeprevir) and ritonavir showed an increased antiviral activity against HCV and was well tolerated by patients (Bourgeois et al., 2017; Vigen et al., 2017).

- Quarfloxin (also known as CX-3543 or itarnafoxin), a fluorooquinolone derivative, is a G-quadruplex-binding compound, which has been clinically evaluated for its antineoplastic activity (Balasubramanian et al., 2011; Drygin et al., 2009). This compound was well tolerated by patients during clinical trials and showed a good toxicity profile, but after reaching Phase II it was discarded for poor bioavailability at the active site (Balasubramanian et al., 2011). Other G-quadruplex binding compounds were shown to exert antiviral activity against different enveloped viruses and to interact with G-quadruplex regions of viral genomes (Lavezzo et al., 2018; Ruggiero and Richter, 2018). For this reason, we hypothesize that quarfloxin could possess a potential antiviral activity, despite to our knowledge it has not yet been evaluated.

### 3.6. Spike glycoproteins sequence alignments

The inhibition mechanism hypothesized in this research strictly depends on the interactions established between the compounds and the HR1 internal residues of the pocket. This mechanism may be consequently extended to other members of the betacoronavirus family depending on the sequence identity found in this region. With this purpose, the sequence of the spike glycoprotein structure used in this study has been first checked against the high number of SARS-CoV-2 sequences deposited until April 1, 2020 in the GenBank Protein database (Clark et al., 2016). We retrieved all complete SARS-CoV-2 spike glycoprotein protein sequences from the NCBI Virus portal (Hatcher et al., 2017), and performed a similarity search using the blast-p suite of BLAST (Altschul et al., 1990), selecting only the region spanning from the fusion peptide to the HR1 region (residues from 811 to 984 in the cryo-EM structure). The 174-residues long query sequence showed 100% identity towards 345 of the 360 evaluated spike-glycoprotein target sequences, indicating that this region is highly conserved between all the SARS-CoV-2 isolates sequenced to date. Among 15 entries with an identity lower than 100%, two sequences showed only one residue mismatch, located outside of the internal cavity, while the others contained undetermined residues that do not match to the query, explaining the lower identity percentage obtained. We further evaluated whether this region was also conserved among other betacoronaviruses (NCBI:txid694002). The submitted query sequence showed a 100% identity with the spike glycoprotein sequence of the bat coronavirus RaTG13 (NCBI:txid2709072; GenBank accession: QHR63300.2). Identity values ranging from about 88 to 95% were observed for different SARS-CoV isolates and other animal SARS-like betacoronaviruses, for a total of 148 different entries. In particular, the internal HR1 region, the specific target of our compounds, is completely conserved among all these 148 hit sequences, except for one residue mismatch observed within the sequence of the palm civet SARS coronavirus A013 (NCBI:txid305402; GenBank accession: AAV97985.1). On the other hand, the spike glycoprotein sequences deposited for the middle east respiratory syndrome coronavirus (MERS-CoV), another concerning human pathogen, only showed about 52–57% sequence identity against the fusion peptide-HR1 SARS-CoV-2 region, and an identity between about 58 and 67% for the internal HR1 region alone.

These results show that our selected compounds, targeting the highly conserved internal regions of the spike, could represent a broad-spectrum treatment active against all SARS-CoV-2 currently isolated variants, overcoming the concerning mutation potential of this virus. Furthermore, the high sequence and structural identities between the SARS-CoV and the SARS-CoV-2 spike glycoproteins suggest that these compounds may represent an efficient therapeutic strategy also towards the SARS-CoV virus. Finally, the low sequence identity evaluated between the SARS-CoV-2 internal HR1 region and that of other members of the betacoronavirus family do not completely exclude the possibility that these drugs may also effectively bind within these non-identical pockets, since the different pattern of interactions that would be established could perhaps enhance the binding of the compounds, instead of preventing it.

### 4. Conclusions

The results obtained in this research support the hypothesis that the presence of specific compounds inside the inner cavity of the spike glycoprotein in prefusion conformation (Fig. 1) could block the HR1 regions, preventing the conformational changes that allow SARS-CoV-2 entry into target cells. In fact, the HR1 regions are actively involved in the compounds recognition, as confirmed by the MM/GBSA per-residue decomposition analysis (Table S2). In addition, our hypothesized inhibitory mechanism may explain the unknown mechanisms by which phthalocyanine derivatives have been experimentally shown to induce viral fusion inhibition with no impairment in spike glycoproteins binding to cell receptors (Zvorot et al., 2003).

We suggest that an optimal therapeutic strategy could involve the simultaneous administration of a cocktail of inhibitors that could boost up each individual antiviral property. Based on the reweighted virtual screening results (Table 3), obtained though the evaluation of the previous literature, we propose the concurrent administration of four compounds: phthalocyanine, hypericin, TMC-647055 and quarfloxin, to be used as a possible antiviral treatment against SARS-CoV-2. Their already observed antiviral properties, although mostly assessed on different targets and viruses, add an important value to these drugs, since they could interfere with SARS-CoV-2 infection also with other mechanisms independent from the hypothesized inhibition of spike glycoprotein fusion to the host cell.

To evaluate the feasibility of the simultaneous localization of different compounds inside the pocket we also performed sequential molecular docking simulations of these four selected drugs (Fig. S4). Results showed that the compounds could fit together inside the pocket still maintaining a high affinity for the HR1 regions and the overall binding pocket, where evaluated interaction energies are: phthalocyanine: -16.7 kcal/mol, hypericin: -13.7, TMC-647055: -12.7 and quarfloxin: -11.6 kcal/mol.

Furthermore, the results obtained carrying out the sequence alignments strongly indicate that the hypothesized inhibition mechanism described in this work may be extended to all the identified SARS-CoV-2 isolates, to the closely related SARS-CoV, and potentially even to other known members of the betacoronavirus family.

Concluding, the hypotheses arising from this simulation research can represent a valid contribution to the international community for the modulation of the SARS-CoV-2 infection, and could also be useful against other enveloped viruses sharing the same infection mechanism.

### Author contributions

M.F., F.I. and A.R. conceived, designed and wrote the paper; A.R. and F.I. performed the calculations and analyzed the data.

### CRediT authorship contribution statement

**Alice Romeo:** Conceptualization, Writing - original draft, Formal analysis, Investigation. **Federico Iacovelli:** Conceptualization, Writing - original draft, Formal analysis, Investigation. **Mattia Falconi:** Project administration, Conceptualization, Methodology, Writing - original draft.
Declaration of Competing Interest

The authors declare they have no potential conflicts of interest to disclose.

Acknowledgments

The computing resources and the related technical support were provided by CRESCO/ENEAGRID High Performance Computing infrastructure (Ponzi et al., 2014). CRESCO/ENEAGRID High Performance Computing infrastructure is funded by ENEA, the Italian National Agency for New Technologies, Energy and Sustainable Economic Development and by Italian and European research programmes, see http://www.cresco.enea.it/english for information.

Appendix A. Supplementary data

Supplementary material related to this article can be found in the online version, at doi:https://doi.org/10.1016/j.virusres.2020.190808.

References

Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., Lindahl, E., 2015. The oﬀ-mass Ewald method. J. Chem. Phys. 142, 134104.https://doi.org/10.1063/1.4914554.

Aoki, K.M., Yoneya, M., Yokoyama, H., 2004. Constant pressure Ms simulation method. J. Comput. Chem. 25 (13), 1605–1612.https://doi.org/10.1002/jcc.20067.

Aoki, K.M., Yoneya, M., Yokoyama, H., 2004. Constant pressure Md simulation method. J. Comput. Chem. 25 (13), 1605–1612.https://doi.org/10.1002/jcc.20067.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215 (3), 403–410.https://doi.org/10.1016/0022-2836(92)90599-E.

Ambraseys, K.A., Fanning, G.C., Lin, A., Blinova, O., Nowroozi, E.P., Ostapchuk, Y., Schäfer, A.A., Brister, J.R., 2017. Virus Variation Resource - improved response to emergent viral outbreaks. Nucleic Acids Res. 45 (D1), D482–D490.https://doi.org/10.1093/nar/gkw1065.

Aoki, K.M., Yoneya, M., Yokoyama, H., 2004. Constant pressure Md simulation method. J. Comput. Chem. 25 (13), 1605–1612.https://doi.org/10.1002/jcc.20067.

Aoki, K.M., Yoneya, M., Yokoyama, H., 2004. Constant pressure Md simulation method. J. Comput. Chem. 25 (13), 1605–1612.https://doi.org/10.1002/jcc.20067.

Aoki, K.M., Yoneya, M., Yokoyama, H., 2004. Constant pressure Md simulation method. J. Comput. Chem. 25 (13), 1605–1612.https://doi.org/10.1002/jcc.20067.

Aoki, K.M., Yoneya, M., Yokoyama, H., 2004. Constant pressure Md simulation method. J. Comput. Chem. 25 (13), 1605–1612.https://doi.org/10.1002/jcc.20067.

Aoki, K.M., Yoneya, M., Yokoyama, H., 2004. Constant pressure Md simulation method. J. Comput. Chem. 25 (13), 1605–1612.https://doi.org/10.1002/jcc.20067.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215 (3), 403–410.https://doi.org/10.1016/0022-2836(92)90599-E.

Aoki, K.M., Yoneya, M., Yokoyama, H., 2004. Constant pressure Md simulation method. J. Comput. Chem. 25 (13), 1605–1612.https://doi.org/10.1002/jcc.20067.

Aoki, K.M., Yoneya, M., Yokoyama, H., 2004. Constant pressure Md simulation method. J. Comput. Chem. 25 (13), 1605–1612.https://doi.org/10.1002/jcc.20067.

Aoki, K.M., Yoneya, M., Yokoyama, H., 2004. Constant pressure Md simulation method. J. Comput. Chem. 25 (13), 1605–1612.https://doi.org/10.1002/jcc.20067.

Aoki, K.M., Yoneya, M., Yokoyama, H., 2004. Constant pressure Md simulation method. J. Comput. Chem. 25 (13), 1605–1612.https://doi.org/10.1002/jcc.20067.

Aoki, K.M., Yoneya, M., Yokoyama, H., 2004. Constant pressure Md simulation method. J. Comput. Chem. 25 (13), 1605–1612.https://doi.org/10.1002/jcc.20067.

Aoki, K.M., Yoneya, M., Yokoyama, H., 2004. Constant pressure Md simulation method. J. Comput. Chem. 25 (13), 1605–1612.https://doi.org/10.1002/jcc.20067.
