Modeling SARS-CoV-2 spike/ACE2 protein–protein interactions for predicting the binding affinity of new spike variants for ACE2, and novel ACE2 structurally related human protein targets, for COVID-19 handling in the 3PM context

Vincenzo Tragni[1] · Francesca Preziusi[1] · Luna Laera[1] · Angelo Onofrio[1] · Ivan Mercurio[1] · Simona Todisco[2] · Mariateresa Volpicella[1] · Anna De Grassi[1,3] · Ciro Leonardo Pierri[1,3]

Abstract

Aims The rapid spread of new SARS-CoV-2 variants has highlighted the crucial role played in the infection by mutations occurring at the SARS-CoV-2 spike receptor binding domain (RBD) in the interactions with the human ACE2 receptor. In this context, it urgently needs to develop new rapid tools for quickly predicting the affinity of ACE2 for the SARS-CoV-2 spike RBD protein variants to be used with the ongoing SARS-CoV-2 genomic sequencing activities in the clinics, aiming to gain clues about the transmissibility and virulence of new variants, to prevent new outbreaks and to quickly estimate the severity of the disease in the context of the 3PM.

Methods In our study, we used a computational pipeline for calculating the interaction energies at the SARS-CoV-2 spike RBD/ACE2 protein–protein interface for a selected group of characterized infectious variants of concern/interest (VoC/ VoI). By using our pipeline, we built 3D comparative models of the SARS-CoV-2 spike RBD/ACE2 protein complexes for the VoC B.1.1.7-United Kingdom (carrying the mutations of concern/interest N501Y, S494P, E484K at the RBD), P.1- Japan/Brazil (RBD mutations: K417T, E484K, N501Y), B.1.351-South Africa (RBD mutations: K417N, E484K, N501Y), B.1.427/B.1.429-California (RBD mutations: L452R), the B.1.141 (RBD mutations: N439K), and the recent B.1.617.1- India (RBD mutations: L452R; E484Q) and the B.1.620 (RBD mutations: S477N; E484K). Then, we used the obtained 3D comparative models of the SARS-CoV-2 spike RBD/ACE2 protein complexes for predicting the interaction energies at the protein–protein interface.

Results Along SARS-CoV-2 mutation database screening and mutation localization analysis, it was ascertained that the most dangerous mutations at VoC/VoI spike proteins are located mainly at three regions of the SARS-CoV-2 spike "boat-shaped" receptor binding motif, on the RBD domain. Notably, the P.1 Japan/Brazil variant present three mutations, K417T, E484K, N501Y, located along the entire receptor binding motif, which apparently determines the highest interaction energy at the SARS-CoV-2 spike RBD/ACE2 protein–protein interface, among those calculated. Conversely, it was also observed that the replacement of a single acidic/hydrophilic residue with a basic residue (E484K or N439K) at the "stern" or "bow" regions, of

the boat-shaped receptor binding motif on the RBD, appears to determine an interaction energy with ACE2 receptor higher than that observed with single mutations occurring at the "hull" region or with other multiple mutants. In addition, our pipeline allowed searching for ACE2 structurally related proteins, i.e., THOP1 and NLN, which deserve to be investigated for their possible involvement in interactions with the SARS-CoV-2 spike protein, in those tissues showing a low expression of ACE2, or as a novel receptor for future spike variants. A freely available web-tool for the in silico calculation of the interaction energy at the SARS-CoV-2 spike RBD/ACE2 protein–protein interface, starting from the sequences of the investigated spike and/or ACE2 variants, was made available for the scientific community at: https://www.mitoairm.it/covid19affinities.
Conclusion In the context of the PPPM/3PM, the employment of the described pipeline through the provided webservice, together with the ongoing SARS-CoV-2 genomic sequencing, would help to predict the transmissibility of new variants sequenced from future patients, depending on SARS-CoV-2 genomic sequencing activities and on the specific amino acid replacement and/or on its location on the SARS-CoV-2 spike RBD, to put in play all the possible counteractions for preventing the most deleterious scenarios of new outbreaks, taking into consideration that a greater transmissibility has not to be necessarily related to a more severe manifestation of the disease.

Abbreviations

Å	Ångstrom (1 Å = 1.0×10^{-10} m or 0.1 nm)
3D	Three dimensional
ACE2	Angiotensin-converting enzyme II
AFM-SMFS	Atomic force microscopy-based single-molecule force spectroscopy
ANPEP	Alanyl aminopeptidase
BLI	bio-layer interferometry
BSG	Basigin (also known as CD147)
CD147	Cluster of differentiation 147
CDS	Coding DNA sequence
Chain-ID	Chain identifier
CLEC4G	Glycan-binding receptor of the C-type lectin domain family 4 member G
CLEC4M	Glycan-binding receptor of the C-type lectin domain family 4 member M
CoV	Coronavirus
CTSB	Lysosomal cysteine peptidase B
CTSL	Lysosomal cysteine peptidase L
DPP4	Dipeptidyl peptidase-4
EC50	Half maximal effective concentration
FACS	Fluorescence-activated cell sorter
GTEx	Genotype-tissue expression
H-bonds	Hydrogen bonds
I-TASSER	Iterative Threading ASSEmbly Refinement
IFITM3	Interferon-induced membrane protein 3
K_D	Equilibrium dissociation constant
MSA	Multiple sequence alignment
NLN	Neurolysin
PDB	Protein Data Bank
PIC	Protein Interactions Calculator
PPPM/3PM	Predictive, preventive and personalized medicine
RBD	Receptor binding domain
RBM	Receptor binding motif
RMSD	Root mean square deviation
SARS	Severe acute respiratory syndrome
THOP1	Thimet oligopeptidase
TMPRSS2	Transmembrane serine protease 2
TMPRSS4	Transmembrane serine protease 4
TPM	Transcripts per million
VoC	Variants of concern: a variant for which there is evidence of an increase in transmissibility, more severe disease, significant reduction in neutralization by antibodies generated during previous infection or vaccination, reduced effectiveness of treatments or vaccines, or diagnostic detection failures (https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html)
VoI	Variants of interest: a variant with specific genetic markers that have been associated with changes to receptor binding, reduced neutralization by antibodies generated against previous infection or vaccination, reduced efficacy of treatments, potential diagnostic impact, or predicted increase in transmissibility or disease severity (https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html#Interest)

Introduction

Covid-19 and mechanism of infection

A coronavirus identified in 2019, SARS-CoV-2, has caused a pandemic of respiratory disease called COVID-19 [1–3]. WHO reports that there have been 244,385,444 confirmed cases of COVID-19, including 4,961,489 deaths (27 October 2021, https://covid19.who.int/) due to the spread of several SARS-CoV-2 variants [1, 4–6]. Although the mechanism of infection was deeply investigated taking to the observation that a direct interaction between the SARS-CoV-2 spike protein and the human ACE2 receptor is responsible for the host-cell virus entry [7–11], a definitive cure for COVID-19 is not available yet [12–16]. Nevertheless, the employment of several approved vaccines [17–20] is helping in preventing the worst scenario for the current outbreak by limiting the most severe effects of the disease in vaccinated people and by reducing virus circulation [21]. Also, the intense SARS-CoV-2 genomic sequencing activity in clinics is helping in the early recognition of new variants that deserve to be studied for evaluating their improved ability in penetrating host-cells and/or in escaping vaccine induced antibodies [22–25].

Along genome sequencing analysis it was observed that variants of concern/interest (VoC/VoI) showing mutations at the SARS-CoV-2 spike receptor binding domain (RBD) are those that need to be monitored more carefully due to their involvement in direct interactions with the human dipeptidyl carboxydipeptidase angiotensin-converting enzyme II (ACE2) receptor, causing the host-cell invasion [7–9].

SARS-CoV-2 spike VoC/VoI

SARS-CoV-2 genomic sequencing activities are crucial for the early detection of new variants and several databases provide clues about VoC and VoI mutations, although it happens that the same mutations can be detected in different variants [4, 26]. Sometimes nomenclature problems and mutation detection in different strains just depend on the mutation prevalence in the lineage, other times it depends on technical annotation problems [4, 26]. Although the effort for sequencing activities and database sequence collection was enormous, it still lacks a quantifiable parameter to be monitored, maybe through an automated in silico tool, for quickly predicting the transmissibility (how easy viruses spread [27]) and virulence (how harmful is a virus for its host [28]) of new variants.

Indeed, at the moment only the direct observation and quantification of a SARS-CoV-2 variant spread [29] together with relatively "slow/controversial" in vitro binding assays [30] provide clues about the transmissibility [27] or

virulence/pathogenicity [28] of new variants, when the virus is already circulating. Some computational tools are used to calculate the stability and/or preliminary immunogenic properties of SARS-CoV-2 mutated proteins [31, 32], but those tools are further time-consuming and cannot be easily adapted in the short-term period for dealing with the current pandemic in the context of the 3PM approaches.

Thus, it might be crucial to identify a parameter to be quickly computationally calculated/quantified, which may correlate with the transmissibility [27] or the virulence [28] of a variant. It should be mentioned that a greater virulence is related to a greater ability of the virus to replicate in the host organism often causing a more severe disease, although resulting usually in a lower transmissibility [33]. Conversely, a lower virulence is often correlated with a higher transmissibility [27], as already observed for recent variants showing new mutations, i.e., D614G, at the SARS-CoV-2 spike protein (relatively far from the RBD region involved in direct interactions with the ACE2 receptor), which resulted in a less deadly virus. As a consequence of the lower lethality, the spike D614G mutation is currently detectable in most of SARS-CoV-2 variants of concern (https://covdb.stanford.edu/page/mutation-viewer; [34–36]).

The controversial role of ACE2 variants in SARS-CoV-2 infection and variant spread

There is also the need to clarify the role of ACE2 variants, existing in the world population (https://gnomad.broadinstitute.org/gene/ENSG00000130234?dataset=exac), in variability and susceptibility to SARS-CoV-2 infection [37–39]. In order to gain new insights about ACE2 role in SARS-CoV-2 infection, it would be useful to sequence directly ACE2 with SARS-CoV-2 genes from hospitalized patients. Indeed, it is well established that protein–protein interactions depend both on residues at the protein–protein interface but also on residues far from the protein–protein interface that can participate to protein folding through long-range interactions [40–44]. Similarly, it is expected that the combination SARS-CoV-2 spike variants interacting with ACE2 variants will produce different interaction energies that may reflect a different susceptibility to COVID-19 [37–39], despite of the average transmissibility and virulence of a strain estimated without taking into account the personal patient genetic context.

More in general, we expect that the calculation of the interaction energies between a specific SARS-CoV-2 spike variant and a specific ACE2 variant, as sequenced from the affected patients, may provide a more robust prediction about the specific SARS-CoV-2 spike/ACE2 interaction energies in a patient or in an asymptomatic/paucisymptomatic carrier.

A modular molecular framework for predicting *in silico* the binding affinity of ACE2 variants for SARS-CoV-2 spike VoC to get preliminary indications about new variant transmissibility and virulence in the context of the 3PM approaches.

In light of what was reported about the role of SARS-CoV-2 spike variants and ACE2 variants, we keep in mind that it would be useful to consider the SARS-CoV-2 spike/ACE2 interaction energy as a quantifiable parameter for predicting SARS-CoV-2 spike/ACE2 binding affinity, which might show a positive correlation with virus virulence (i.e., the ability of the virus to enter host-cells and consequently replicate itself, resulting in more harmful or even deadly consequences [28]) and a negative correlation with virus transmissibility (i.e., the ability of the virus to spread in the population causing a less severe disease [27, 33]). The calculation of the interaction energy at the SARS-CoV-2 spike RBD/ACE2 protein–protein interface, would help to get a preliminary indication about the possible increased/decreased virulence of a new strain. Indeed, we would expect that an increased SARS-CoV-2 spike/ACE2 binding affinity might reflect an increased ability of the virus to replicate within the host organism with more severe outcomes, whereas a decreased binding affinity may result in a less lethal but more transmissible virus [27, 33], as already observed in the correlation between protein–protein binding affinity along pathogen–host interactions and pathogen virulence in other biological systems [45–47].

With this aim, we develop a computational pipeline to calculate the interaction energies at the SARS-CoV-2 spike/ACE2 protein–protein interface, starting from the SARS-CoV-2 spike and ACE2 amino acid sequences, as sequenced from patients, to predict the binding affinity of the spike protein of new variants for a specific human ACE2 variant.

The computational pipeline here presented consists of a modular molecular framework that was initially used for investigating SARS-CoV-2 spike receptor binding domain (RBD)/ACE2 interactions and the spike pre/post-fusion conformational changes [7].

Now, it was modified to quickly predict binding affinities of the human ACE2 receptor for the spike protein RBD of infectious VoC/VoI, as long as SARS-CoV-2 genome sequencing activities quickly indicate the existence of new variants. Indeed, by using our pipeline and starting from sequencing data, it is possible to build a reliable 3D model of the SARS-CoV-2 spike protein in complex with ACE2 for the following calculation of the interaction energy at the SARS-CoV-2 spike RBD/ACE2 protein–protein interface and for the estimation of their binding affinity, to be compared with the interaction energies calculated between the hCoV.19Wuhan.WIV04.2019 spike sequence and/or other characterized spike variants (i.e., those proposed in this manuscript) and the reference ACE2 protein [7].

Thus, aiming to test our pipeline and to investigate the effect of amino acid replacement at the SARS-CoV-2 spike RBD, as observed within the most studied VoC, we built a 3D comparative model for each spike variant showing an amino acid replacement at the RBD positions K417; N439; L452; E484; S477; S494; N501, as highlighted in the SARS-CoV-2 VoC B.1.1.7-United Kingdom (UK, mutations of concern/interest at the RBD: N501Y, S494P, E484K), P.1-Japan/Brazil (K417T, E484K, N501Y), B.1.351-South Africa (S. Africa, mutations of concern/interest: K417N, E484K, N501Y), B.1.427/B.1.429-California (mutations of concern/interest: L452R), the B.1.141 variant (mutations of concern/interest: N439K), the recent B.1.617.1-India (mutations of concern/interest: L452R; E484Q), and the B.1.620 (mutations of concern/interest: namely S477N; E484K). Notably, all the cited VoC genomes have been sequenced in the last year and are responsible for the evolution of current pandemics situation [6, 25, 48–52].

Knowing the binding affinity of the SARS-CoV-2 spike/ACE2 protein complexes in case of detection of new SARS-CoV-2 spike variants might be crucial to gain clues about transmissibility and virulence of the new detected variant. Indeed, it is recognized that a greater spike/ACE2 affinity may result in a more rapid spread of the investigated variant and in the loss of efficiency of vaccines based on a specific spike elder sequence, i.e., the YP_009724390.1 sequence and the related coding DNA sequence (CDS), namely the hCoV.19Wuhan.WIV04.2019 sequence, https://www.gisaid.org/hcov19-variants/) [19, 20].

On the other hand, a similar approach can be trained by coupling the predicted binding affinity values with the number of infected people and with data about disease severity (from the known variants and/or from current and future hospitalized patients) and the trained algorithm can be used to model the interaction between next SARS-CoV-2 spike variants and ACE2 variants to predict their binding affinities for improving transmissibility, virulence and disease severity prediction, for the early recognition of the most dangerous variants.

This knowledge might help to early alert the scientific community to think about new rounds of vaccination by using new vaccines based on a cocktail of the newly identified (or predicted as dangerous) spike sequences (in the form of mRNA [19, 20] or proteins/subdomains [53–55]) showing high affinities for the human host-cell receptors and/or to design new targeted antibodies [7, 56] for preventing new outbreaks due to the more virulent variants.

Other possible cell-entry factors

Finally, given the low expression of ACE2 observed in some tissues highly exposed to SARS-CoV-2, and the acquired knowledge about a set of possible human host-cell entry

factors [57–60], we used our computational pipeline for predicting ACE2 structurally related receptors, aiming to identify other putative SARS-CoV-2 entry sites in human host cells, based on a folding recognition approach, more than on sequence/functional comparative analysis.

Materials and methods

Comparative 3D modelling of SARS-CoV-2 spike RBD and the investigated RBD mutants interacting with the human ACE2 receptor

Starting from the atomic coordinates of the SARS-CoV-2 spike RBD protein domain extracted from 6m0j.pdb (according to the YP_009724390.1 sequence), we built the 3D comparative models of the investigated RBD mutants with specific reference to the single mutants N501Y, E484K/Q, N439K, K417N/T, L452R, S477N, and S494P; the double mutants L452R-E484Q and S477N-E484K; and the triple mutants N501Y-E484K-K417N, N501Y-E484K-S494P, and N501Y-E484K-K417T by using the mutagenesis tool implemented in SwissPDBViewer [61].

In order to obtain a pose of the 3D protein complex of SARS-CoV-2 spike RBD mutants interacting with ACE2, the built 3D comparative models of the mutants were superimposed to the 3D protein complex consisting of the Wuhan SARS-CoV-2 spike RBD interacting with ACE2, available under the 6m0j.pdb protein data bank (PDB) entry.

All the generated 3D all-atom models were energetically minimized using the Yasara Minimization server and residues packing was checked and repaired, where necessary, according to the FOLDX repair function [62]. PyMol [63] (https://www.pymol.org) was then used for examining (by manual inspection) the obtained 3D structure models, and for checking the correct packing of local secondary structures.

The variation in the number of interactions (hydrogen bonds (H-bonds), ionic and aromatic interactions) at the SARS-CoV-2 spike RBD/ACE2 interface in presence of the investigated amino acid replacements have been calculated by using the PIC webserver [64] and verified by manual inspection by using PyMOL.

Crystal structure sampling of possible ACE2 structural related alternative receptors *via* folding recognition and multiple sequence alignments (MSAs)

ACE2 structural related proteins were sampled by using the folding recognition methods implemented in pGenThreader [65] and I-Tasser [66]. With this aim, the amino acid sequence of ACE2 (NP_001358344.1) was used as query

sequence for running pGenThreader (http://bioinf.cs.ucl.ac.uk/psipred/) and I-Tasser (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) to screen the PDB, searching for ACE2 structurally related crystallized proteins [7, 65–71].

The crystallized structures of the proteins sampled by pGenThreader and I-Tasser analysis were structurally aligned with the 3D coordinates of ACE2 available under the PDB_ID 6m0j.pdb. For obtaining the structural alignment, we used the "super" command available in Pymol [63], which is able to structurally align also proteins with a lower percentage of identical residues, due to its ability in providing a sequence-independent structure-based pairwise alignment [7, 67, 69, 70].

The sequences of the investigated crystallized ACE2 structurally related proteins were aligned using ClustaW [72] and optimized by visual inspection based on the structural alignment obtained by PyMOL [63].

Preparation of 3D complex protein models hosting the Wuhan SARS-CoV-2 spike RBD in complex with the highlighted ACE2 structurally related receptors

In order to obtain the most likely 3D protein complexes of SARS-CoV-2 spike RBD interacting with the highlighted ACE2 structurally related receptors "dipeptidyl carboxydipeptidase angiotensin I converting enzyme 1 (ACE, https://www.ncbi.nlm.nih.gov/gene/1636)," "thymet oligopeptidase 1 (THOP1, https://www.ncbi.nlm.nih.gov/gene/7064)," and "neurolysin peptidase (NLN, https://www.ncbi.nlm.nih.gov/gene/57486)," the sampled crystallized structures of ACE, THOP1 and NLN were superimposed on the structure of the human ACE2 receptor within 6m0j.pdb, crystallized in complex with the SARS-CoV-2 spike RBD.

The obtained 3D coordinates of SARS-CoV-2 spike RBD (according to the YP_009724390.1 sequence as obtained from 6m0j.pdb) in complex with ACE, THOP1, and NLN were each saved in a new PDB file.

The variation in the number of interactions (hydrogen bonds (H-bonds), ionic and aromatic interactions) at the SARS-CoV-2 spike RBD/ACE2 structurally related proteins (i.e., RBD/ACE; RBD/THOP1; RBD/NLN) interface have been calculated by using the PIC webserver [64] and verified by manual inspection by using PyMOL.

FoldX energy calculations

The 3D coordinates of the Wuhan SARS-CoV-2 spike RBD crystallized in complex with the human ACE2 receptor available under the PDB_ID 6m0j.pdb were used to estimate the binding affinity of the SARS-CoV-2 spike RBD for the human ACE2 (i.e., from the reference sequence NP_001358344.1) by using the FoldX

AnalyseComplex assay [73]. The calculated binding affinity of the two protein domains within the crystallized 6m0j.pdb was used as a reference value for the following comparative analyses [7].

Indeed, the FoldX AnalyseComplex assay was performed to determine the interaction energy between the investigated minimized protein complexes consisting of the SARS-CoV-2 spike RBD mutants (N501Y, E484K/Q, N439K, K417N/T, L452R, S477N, S494P, the double mutants L452R-E484Q, S477N-E484K, and the triple mutants N501Y-E484K-K417N, N501Y-E484K-S494P, and N501Y-E484K-K417T) and ACE2.

Furthermore, the FoldX AnalyseComplex assay was used for determining the binding affinity between SARS-CoV-2 spike RBD and the ACE2 structurally related receptors ACE, THOP1, and NLN, identified through pGenTHREADER and/or I-TASSER analyses.

The way the FoldX AnalyseComplex operates is by unfolding the selected targets and determining the stability of the remaining molecules and then subtracting the sum of the individual energies from global energy [7]. More negative energy values indicate a better binding, whereas positive energy values indicate no binding [73, 74].

GTEx expression map of human coronavirus entry factors

The GTEx portal (https://gtexportal.org/home/, [75]) was screened to assess the gene expression levels of ACE2 structurally related receptors, i.e., ACE, THOP1, and NLN, and of many other genes proposed to play a crucial role in host-cell virus entry processes, i.e., the immunoglobulin basigin (BSG, also known as CD147, https://www.ncbi. nlm.nih.gov/gene/682), the interferon-induced membrane protein IFITM3 (known for being involved in the protection against several viruses, https://www.ncbi.nlm.nih. gov/gene/10410), the FURIN protease (https://www.ncbi. nlm.nih.gov/gene/5045), the lysosomal cysteine peptidases CTSB (https://www.ncbi.nlm.nih.gov/gene/1508) and CTSL (https://www.ncbi.nlm.nih.gov/gene/1514), the membrane alanyl aminopeptidase ANPEP (https://www. ncbi.nlm.nih.gov/gene/290), the transmembrane serine proteases TMPRSS2 (https://www.ncbi.nlm.nih.gov/gene/ 7113) and TMPRSS4 (https://www.ncbi.nlm.nih.gov/gene/ 56649), the dipeptidyl peptidase DPP4 (also known as CD26, https://www.ncbi.nlm.nih.gov/gene/1803), the glycan-binding receptors of the C-type lectin family (known for being involved in the recognition of several viruses and bacteria) CLEC4G (https://www.ncbi.nlm.nih.gov/ gene/339390) and CLEC4M (https://www.ncbi.nlm.nih. gov/gene/10332), and the lymphocyte antigen 6E LY6E (https://www.ncbi.nlm.nih.gov/gene/4061) [59, 60].

Results

RBD amino acid replacements, number of interactions, and local secondary structure perturbation

The analysis about SARS-CoV-2 VoC spread in the last year has highlighted the important role played by amino acid replacements occurring at the spike RBD in SARS-CoV-2 host-cell entry and virus infection [7–9, 11, 18, 22–25, 76]. Here, we report about the employment of a computational pipeline for the estimation of binding affinity and interactions between SARS-CoV-2 spike RBD variants and ACE2 in presence of mutations occurring at seven positions of the spike RBD, as observed in seven VoC (Fig. 1), chosen among the best characterized variants (https://www.ecdc. europa.eu/en/covid-19/variants-concern).

The replacement of the seven investigated residues (N501Y, yellow sticks; E484K/Q, green/pink sticks; N439K, orange sticks; K417N/T, cyan/teal sticks; L452R, dark-blue sticks; S477N, light pink sticks; S494P, hot-pink sticks; Fig. 2) introduces an important perturbation in the local secondary structure of the "boat-shaped" receptor binding motif (RBM) located on the RBD head, consisting of a "bow" portion (residues 456–459 and 468–490, 6m0j. pdb RBD sequence numbering, Fig. 2), a "hull" portion (residues 450–455 and 491–496, 6m0j.pdb RBD sequence numbering, Fig. 2), and a "stern" portion (residues 436–449 and 497–503, 6m0j.pdb RBD sequence numbering, Fig. 2). Notably, the RBM highlighted on the head of the spike RBD is the main RBD portion involved in direct binding interactions with ACE2, mainly consisting of H-bonds and/or ionic and aromatic interactions.

More in detail, the investigated L452R and E484Q amino acid replacements cause a local re-arrangement that perturbs the small beta-sheet in the "hull" region of the boat-shaped RBM (residues 450–455, Fig. 2), whereas N501Y, K417N, E484K and N439K amino acid replacements cause a conformational change in an alpha-helix close to N501 and N439 in the "stern" region (residues 436–449 and 497–503). Although S494P, approximately located in the center of the "boat-hull" region (residues 491–496), and S477N, located on the tip of the "boat-bow" region (residues 468–490) do not produce a viewable local conformational change, it is well known that the replacement of a serine may confer a different flexibility to the local secondary structure elements hosting the investigated mutation, due to the different abilities of Ser/Pro/Asn/Thr residues in producing kink/hinge movements [77, 78].

The perturbation of the local secondary structure of structural elements hosting the cited amino acid
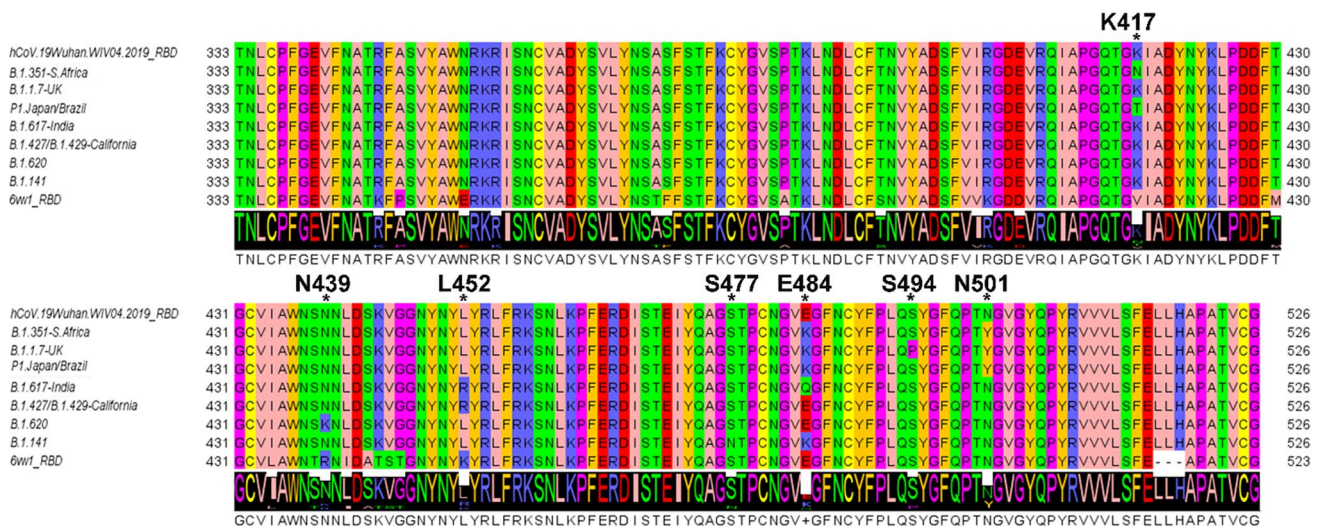
Fig. 1 Multiple sequence alignment (MSA) of SARS-CoV-2 spike RBD amino acid sequences highlighted from the reported sequenced VoC**.** MSA of the SARS-CoV-2 spike RBD sequence from the crystallized SARS-CoV-2 spike protein (PDB_ID: 6m0j.pdb, according to the Wuhan YP_009724390.1 sequence), and the spike RBDs as sequenced from the reported VoC, showing at least an amino acid replacement at the RBD positions K417; N439; L452, E484; S494; N501. The "*" symbols and the labels indicate the SARS-CoV-2 spike RBD positions involved in an amino acid replacement. The sequence of the SARS-CoV-2 chimeric RBD (showing a three amino acid deletion and 22 missense mutations, not detected in the investigated VoC) from 6vw1.pdb is reported for comparative purposes. The MSA is colored according to the JalView Zappo style (green: hydrophilic residues (N,S,Q,T); salmon: aliphatic/hydrophobic residues (V, I,L, A); orange aromatic residues (Y, F, W); yellow: cysteine residues; magenta: conformationally special residues (P, G); red: acidic negative residues (D, E); blue: basic positive residues (R, K), i.e., see: http://www.jalview.org/help/html/colourSchemes/zappo.html)

replacements also triggers the formation of new H-bonds, ionic and aromatic interactions and local conformational changes at the protein–protein interface along interactions with ACE2 (Figs. 3 and 4, Table 1, Supplementary Table 1).

Indeed, it is possible to count a slight increase in the hydrophobic interactions within a range of 5 Å at the SARS-CoV-2 spike RBD/ACE2 interface in the B.1.1.7-UK (S494P_N501Y_E484K); in the P.1 Japan/Brazil (N501Y_E484K_K417T) or in the B.1.351 S. Africa (N501Y_E484K_K417N) VoC (Supplementary Table 1 and Fig. 4).

Some variations are observed in the number of protein–protein main chain–side chain hydrogen bonds which are apparently decreased at the SARS-CoV-2 spike RBD/ACE2 interface in the B.1.1.7-UK (S494P_N501Y_E484K), in the B.1.617.1-India (E484Q_L452), in the B.1.427/B.1.429 California (L452R), and in the B.1.617. Indian (E484Q_L452R) VoC, whereas the same interactions appear increased in number in the P.1 Japan/Brazil (N501Y_E484K_K417T) and in the B.1.141 VoC N439K (Supplementary Table 1 and Fig. 4).

A larger number of variations are observed in the number of protein–protein side chain–side chain hydrogen bonds. Indeed, side chain–side chain H-bonds decrease at the SARS-CoV-2 spike RBD/ACE2 interface in all the VoC with the exception of the B.1.427/B.1.429 California L452R VoC (Supplementary Table 1 and Fig. 4).

New protein–protein aromatic–aromatic interactions (within 4.5 and 7 Å) and protein–protein cation-π interactions are observed at the SARS-CoV-2 spike RBD/ACE2 interface in all the VoCs showing the N501Y amino acid replacement (with specific reference to B.1.1.7-UK (S494P_N501Y_E484K), P.1 Japan/Brazil (N501Y_E484K_K417T) and B.1.351 S. Africa (N501Y_E484K_K417N) showing new aromatic interactions with the ACE2 Tyr41 and cation-π interactions with the ACE2 Lys353 (Supplementary Table 1 and Fig. 4). The number of protein–protein ionic interactions (within 6 Å) at the SARS-CoV-2 spike RBD/ACE2 interface may increase depending on the investigated amino acid replacement and its steric hinderance. That is, L452R is the mutant with the greater number of protein–protein ionic interactions (Supplementary Table 1 and Fig. 4).

Calculation of the interaction energy at the SARS-CoV-2 spike RBD/ACE2 protein–protein interface

From an energetical point of view, the P.1 Japan/Brazil VoC, showing the three mutations N501Y_E484K_K417T at the RBD, has the highest binding affinity (-21.37 kcal/mol or -89.41 kJ/mol; Table 1) for ACE2 (increased of 4% with respect to the Wuhan spike RBD, -20.51 kcal/mol or -85.81 kJ/mol, Table 1), followed by the B.1.141-VoC showing the single N439K amino acid replacement at the
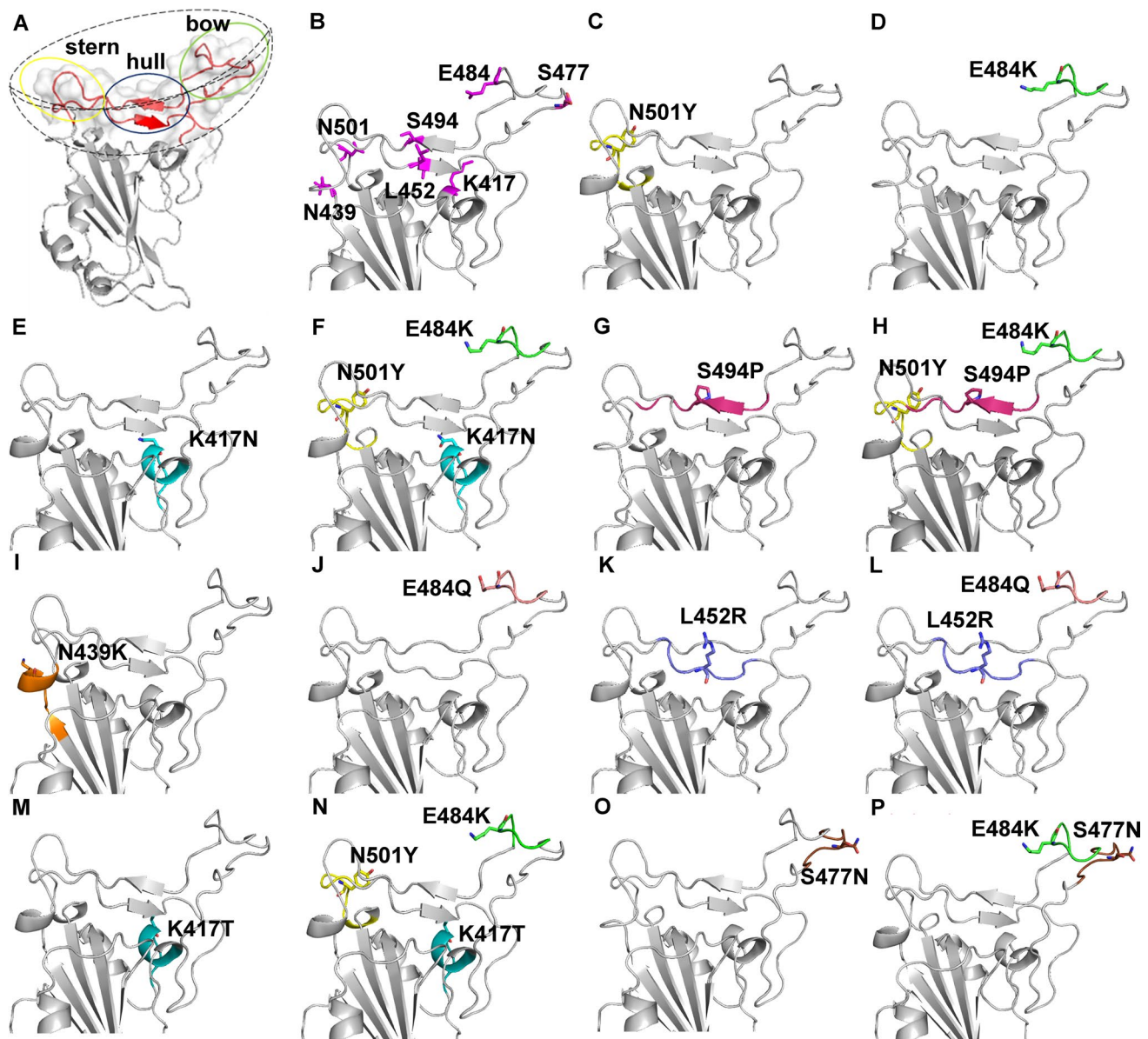
Fig. 2 3D comparative models of the investigated RBD mutants. **A** The SARS-CoV-2 spike RBD crystallized structure (6m0j.pdb) is reported in white cartoon representation. The RBM is highlighted in red cartoon and transparent surf representation. A boat-shaped dashed line is reported around the RBM for defining the 3 different regions, namely the "stern" region (yellow circle), consisting of residues 436–449 and 497–503, the "hull" region (blue circle), consisting of residues 450–455 and 491–496, and the "bow" region (green circle), consisting of residues 456–459 and 468–490 (6m0j.pdb, RBD sequence numbering). Those three regions delimit the RBM area directly involved in binding interactions with ACE2. All the investigated VoC show missense mutations at the 3 highlighted regions within the RBM. **B** Zoomed view of the RBD (white cartoon) showing the RBD amino acids K417; N439; L452; S477; E484; S494; N501 (magenta sticks) observed mutated in the investigated VoC. **C–N** Zoomed views of the RBD investigated mutants. **C** N501Y, yellow sticks; **D** E484K, green sticks; **E** K417N cyan sticks, **F** the triple mutant N501Y_E484K_K417N, observed in the B.1.351 S. Africa VoC; **G** S494P in dark pink sticks; **H** the triple mutant E484K_S494P_N501Y detected in the B.1.1.7_UK variant; **I** N439K, orange sticks, observed in the B.1.141 VoC; **J** E484Q, light pink sticks; **K** L452R in dark-blue sticks, observed in the B.1.427/B.1.429 California VoC; **L** the double mutant E484Q_L452R observed in the B.1.617.1 India VoC; **M** K417T in teal sticks; **N** the triple mutant K417T_E484K_N501Y observed in the P.1 Japan/Brazil VoC; **O** S477N, brown sticks; **P** the two RBD mutations S477N_E484K, observed in the B.1.620 VoC

RBD and the single mutant E484K firstly detected in the spike RBD of the P.1 Japan/Brazil VoC (Table 1). All the other single and multiple amino acid replacements show a slightly decreased interaction energy with ACE2 (Table 1), at variance with what observed for the Wuhan SARS-CoV-2 spike RBD/ACE2 interactions.

Fig. 3 SARS-CoV-2 spike RBD–ACE2 protein–protein binding interactions in presence of the investigated amino acid replacements. **A** The zoomed view of ACE2 (yellow cartoon) interactions with SARS-CoV-2 spike RBD head (white cartoon) is reported. Amino acid positions involved in the investigated mutants are reported in magenta sticks. **B–G** Zoomed views of ACE2 interact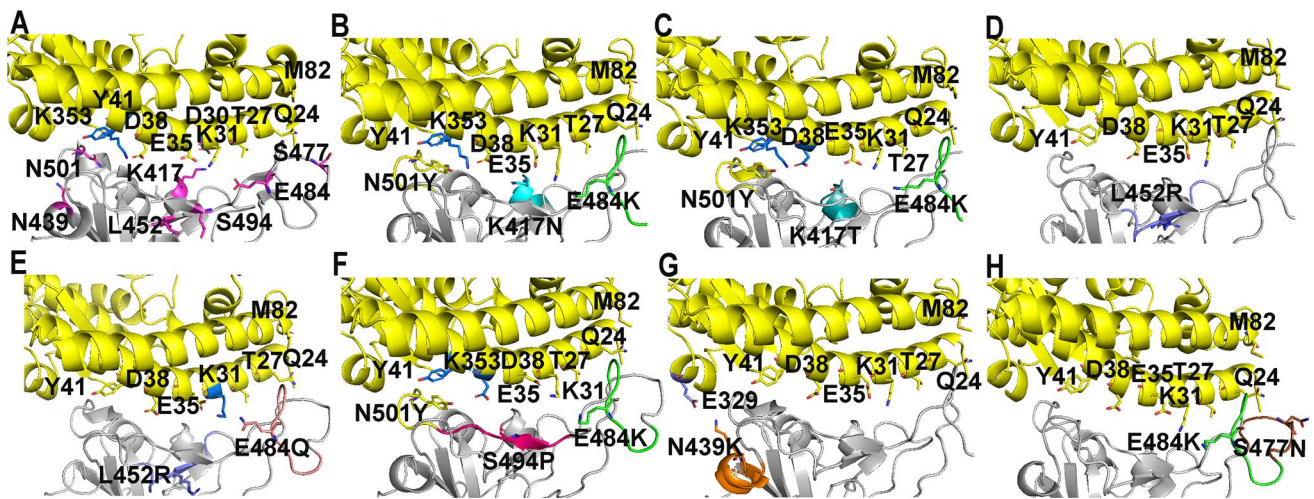ions with the SARS-CoV-2 spike RBD head. **B** B.1.351 S. Africa VoC. **C** P.1 Japan/ Brazil VoC. **D** B.1.427/B.1.429 California VoC. **E** B.1.617.1 India VoC. **F** B.1.1.7-UK variant. **G** B.1.141 VoC. **H** B.1.620 VoC. ACE2 is reported in yellow cartoon in all the panels. ACE2 residues within 4 Å from the RBD are reported in yellow sticks, whereas ACE2 residues within 4 Å from the investigated RBD mutations are reported in blue sticks in all the panels

With the exception of the B.1.427/B.1.429 California VoC (L452R) and B.1.141 (N439K) showing one more interaction (39 interactions) with respect to the interactions detected at the Wuhan SARS-CoV-2 spike RBD/ACE2 protein–protein interface (38 residues), all the other investigated ACE2/RBD complexes showed a decrease in the number of detected interactions at the protein–protein interface caused by the investigated amino acid replacements. The B.1.1.7-UK (S494P_N501Y_E484K) variant and the B.1.617-India (E484Q_L452R) VoC show the lowest number of interactions (27 and 29, respectively) at the RBD/ACE2 protein–protein interface, according to PIC estimations (Fig. 4 and Supplementary Table 1).

A webservice for predicting the SARS-CoV-2 spike RBD/ACE2 binding affinity for gaining clues about transmissibility and virulence

In order to allow clinicians and researchers to predict the binding affinity of new SARS-CoV-2 spike variants for ACE2 variants through our modular molecular framework and pipeline, in the context of the 3PM approaches, a webservice has been made available from our laboratory at the following link https://www.mitoairm.it/covid19affinities. Clinicians and researchers can directly upload through the provided link both the sequenced spike RBD and ACE2 amino acid sequences for calculating the SARS-CoV-2 spike RBD/ACE2 interaction energies to be compared with interaction energies obtained at the protein–protein interface within the crystallized protein complex (6m0j.

pdb) consisting of the hCoV.19Wuhan.WIV04.2019 RBD (YP_009724390.1) and the ACE2 reference protein (NP_001358344.1) and/or the other interaction energies data available from this manuscript, for estimating a possible higher transmissibility or virulence of future detected variants, depending on an increased predicted SARS-CoV-2 spike RBD/ACE2 binding affinity.

Sampling of ACE2 structurally related alternative receptors via folding recognition and multiple sequence alignments (MSA)

ACE2 sequence was used for screening the PDB, searching for ACE2 structurally related proteins that might work as host-cell entry sites for SARS-CoV-2. The performed screening revealed the presence of several crystallized structures of ACE2 and of the structural/functional related ACE protein from several mammalia and insecta species (Table 2). Furthermore, the screening revealed the presence of several oligopeptidases from bacteria and protista, structurally related to ACE2 in the PDB (Supplementary Table 2). Notably, two mammalia oligopeptidases, THOP1 and NLN, were highlighted as ACE2 structurally related proteins (Table 2). ACE2 [79] shares with ACE [80], THOP1 [81], and NLN [82] the 34% (49%), 18% (33%), and 17% (35%) of identical (similar) residues.

The four proteins share a very similar overall structure (Fig. 5). Indeed, the root mean square deviation (RMSD) of the atomic coordinates of the available THOP1 and NLN crystallized structures and the
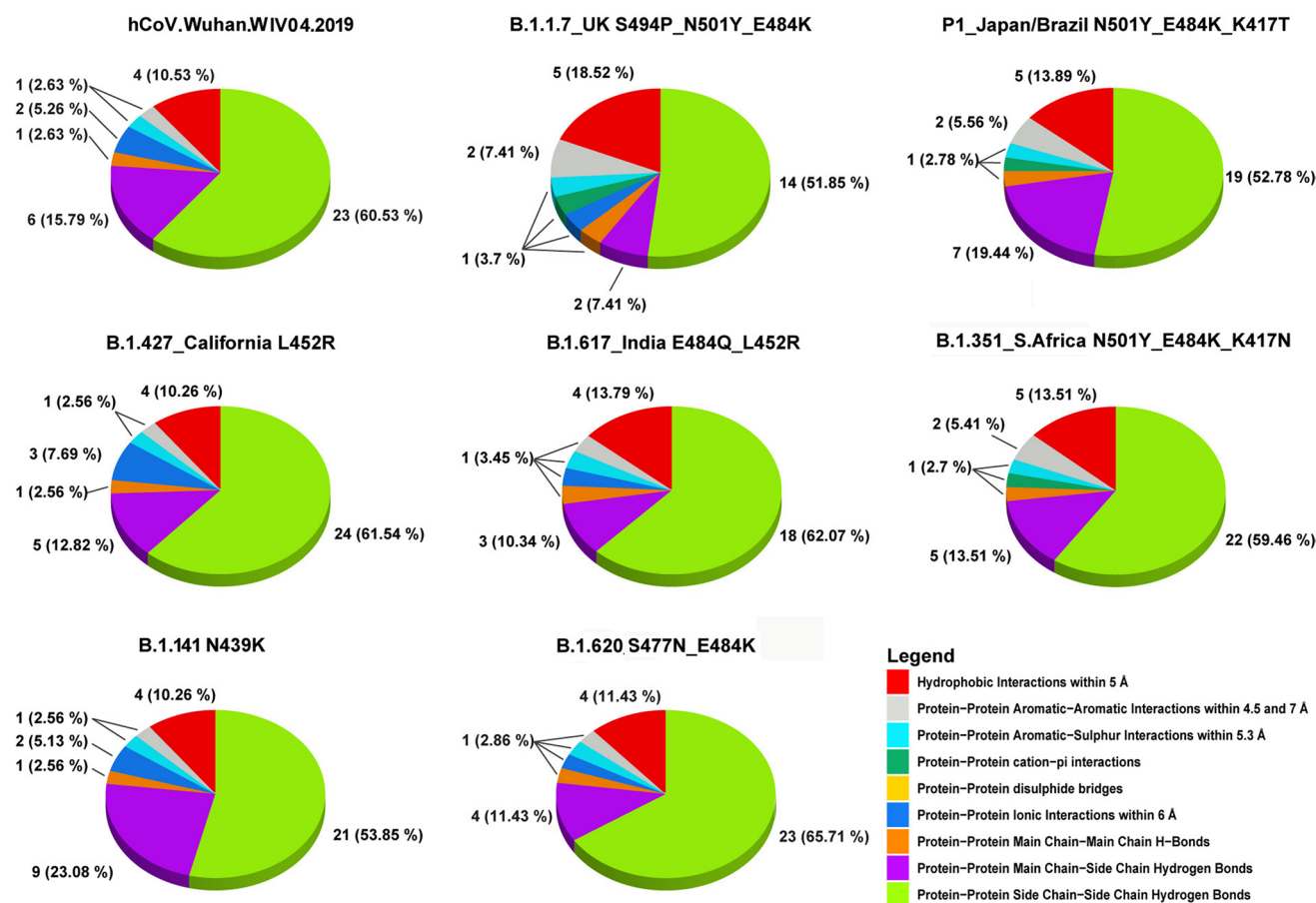
Fig. 4 Pie charts summarize the SARS-CoV-2 spike RBD/ACE2 interactions detected by PIC analysis. The pie charts are realized with R (version 4.0.5) by means of ggplot2 library. For each pie chart, labels are used for indicating the investigated VoC and the number of interactions detected at the SARS-CoV-2 spike RBD/ACE2 protein–protein interface for each interaction type/color reported in the legend. The corresponding percentage of a specific interaction type among the total counted interactions is reported between brackets. It is possible to see that H-bonds are the most represented interactions. Variations in the size of the different pie-chart portions reflect a change in the detected interactions at the SARS-CoV-2 spike RBD/ACE2 protein–protein interface as a consequence of the indicated mutations. By comparing the reported number of interactions with the calculated binding energies (Table 1), it might be speculated that mutations causing a decrease in the "protein–protein side chain-side chain H-bonds" and simultaneously an increase in the "protein–protein main chain–side chain" H-bonds can be responsible for an increase in the interaction energy at the SARS-CoV-2 spike/ RBD protein–protein interface, as a consequence of a greater direct involvement of both protein-backbones in protein–protein interactions

investigated ACE2 structure ranges between 3.9 and 4.5 Å (Table 2, Fig. 5).

SARS-CoV-2 spike RBD variant interactions with the ACE2 structurally related proteins: a comparative analysis (number of interactions, local secondary structure perturbation, and interaction energy calculation)

At the protein–protein interface, it is possible to see that ACE shows two alpha-helices, similarly oriented to the two alpha-helices that represent in ACE2 the main surface of interaction with SARS-CoV-2 spike RBD. THOP1 and NLN show the same alpha-helices in close contact with the SARS-CoV-2 spike RBD and an extra helix parallel to the previous two, forming other interactions with the RBD of the SARS-CoV-2 spike RBD (Fig. 6).

Although the investigated proteins share a low percentage of identical residues, local secondary structures crucial for interactions with the SARS-CoV-2 spike RBD, consisting of the investigated receptor residues located within 4 Å from the RBD are very similar (Fig. 7). The ACE protein region in contact with the RBD is the most similar to the corresponding counterpart in ACE2 (Fig. 7). The four proteins share one longer helix similarly oriented (residues L29-Q50, ACE2 sequence numbering; and residues L65-Q86, THOP1 sequence numbering) and one superimposable beta-sheet different in length (residues T347-L359, ACE2 sequence numbering, and residues A426-L459, THOP1 sequence numbering). THOP1 and NLN show an extra helix consisting of their N-terminal region (residues L24-T64, THOP1 sequence numbering) and show a beta-sheet in place of the helix located

Table 1 FoldX interaction energies between the ACE2 human receptor and the investigated RBD amino acid replacements. Please, note that more negative "interaction energy" values, reported here in kcal/mol, indicate stronger interaction energies and thus higher binding affinity at the indicated SARS-CoV-2 spike RBD/ACE2 protein–protein interface [7, 73]. The energy terms and contributions are reported in kcal/mol according to FOLDX indications. Each energy term has a specific weight in the calculation of the interaction energy, according to the equations reported in [73]. For a complete explanation of item names and energy terms, please visit the link: http://foldxsuite.crg.eu/command/AnalyseComplex. Interaction energies are also reported in kJ/mol (1 kcal = 4.184 kJ). The calculated energy terms and item names are reported in the first column of each row (i.e., in the first cell of the reported table), whereas items (protein chains) and the calculated energy terms are reported for each indicated VoC and mutation within columns 2–16. For a list of all the amino acid abbreviations, please visit https://iupac.qmul.ac.uk/AminoAcid/A2021.html

Items and energy terms	Wuhan SARS-CoV-2 Spike/ACE2 (6m0j.pdb)	N501Y	E484K	K417N	K417T	S494P	E484Q	S477N	B.1.427/B.1.429 California (L452R)	B.1.141 N439K	B.1.351 S. Africa (N501Y_E484K_K417N)	B.1.617.1 India (E484Q_L452R)	P.1 Japan/Brazil (N501Y_E484K_K417T)	B.1.1.7_UK (S494P_N501Y_E484K)	B.1.620 (S477N_E484K)
VoC and mutations															
Group 1 (ACE2)	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Group 2 (RBD)	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E
Interaction energy (kJ/mol)	**−85.81**	−82.0	**−86.48**	−81.08	−76.86	−81.04	−79.62	−79.32	−84.35	**−86.94**	−78.66	−78.91	**−89.41**	−75.81	−76.40
Interaction energy (kcal/mol)	**−20.51**	−19.60	**−20.67**	−19.38	−18.73	−19.37	−19.03	−18.96	−20.16	**−20.78**	−18.80	−18.86	**−21.37**	−18.12	−18.26
Intra-clashes group1 (kcal/mol)	22.29	21.70	25.14	21.20	23.84	22.98	23.69	25.03	24.12	20.69	23.13	24.44	21.94	27.75	24.71
Intra-clashes group2 (kcal/mol)	5.49	4.34	7.10	7.53	5.47	5.51	8.10	5.58	6.60	6.35	7.27	5.73	6.52	5.81	5.57
Backbone H-bond (kcal/mol)	−3.64	−3.01	−3.90	−2.58	−2.71	−2.89	−2.84	−2.80	−2.67	−2.67	−3.08	−3.27	−2.95	−1.72	−2.81
Sidechain H-bond (kcal/mol)	−12.12	−12.85	−12.89	−10.55	−9.39	−7.09	−8.93	−11.03	−10.77	−12.03	−11.57	−8.52	−11.59	−7.09	−10.67
Van der Waals (kcal/mol)	−15.89	−16.63	−16.28	−15.44	−15.59	−15.51	−15.70	−15.74	−15.92	−15.78	−16.54	−15.13	−15.84	−15.64	−16.13

Table 1 (continued)

Items and energy terms	VoC and mutations														
	Wuhan SARS-CoV-2 Spike/ACE2 (6m0j.pdb)	N501Y	E484K	K417N	K417T	S494P	E484Q	S477N	B.1.427/B.1.429 California (L452R)	B.1.141 N439K	B.1.351 S. Africa (N501Y_E484K_K417N)	B.1.617.1 India (E484Q_L452R)	P.1 Japan/Brazil (N501Y_E484K_K417T)	B.1.1.7_UK (S494P_N501Y_E484K)	B.1.620 (S477N_E484K)
Electrostatics (kcal/mol)	−2.41	−2.79	−2.65	−1.82	−1.90	−2.55	−2.42	−2.28	−3.05	−2.50	−1.10	−2.32	−1.11	−2.21	−2.49
Solvation polar (kcal/mol)	22.14	23.39	22.95	20.34	20.79	21.07	20.61	21.90	22.65	21.56	21.33	19.78	20.33	19.91	22.58
Solvation hydrophobic (kcal/mol)	−20.01	−20.93	−20.30	−19.62	−19.81	−19.90	−19.88	−19.80	−19.85	−20.18	−21.33	−19.45	−20.38	−20.56	−20.13
Van der Waals clashes (kcal/mol)	0.41	0.81	0.33	0.36	0.23	0.38	0.57	0.35	0.42	0.42	0.43	0.39	0.45	0.78	0.32
Entropy sidechain (kcal/mol)	9.53	10.72	10.13	8.75	8.84	6.25	8.54	9.78	7.59	9.81	10.56	7.87	9.05	7.43	9.66
Entropy mainchain (kcal/mol)	1.57	1.78	1.70	1.09	0.77	1.04	1.24	0.80	1.73	0.85	2.34	1.67	0.70	1.15	1.68
Torsional clash (kcal/mol)	0.08	0.06	0.42	0.08	0.08	0.07	0.08	0.09	0.06	0.05	0.17	0.37	0.05	0.10	0.07
Backbone clash (kcal/mol)	2.49	2.20	2.57	1.84	2.23	2.20	2.23	2.25	2.14	2.49	2.33	2.18	2.30	2.00	2.24
Helix dipole (kcal/mol)	−0.06	−0.04	0.01	−0.10	−0.11	−0.06	−0.09	−0.12	−0.10	−0.02	−0.04	−0.09	−0.12	−0.03	−0.12
Disulfide (kcal/mol)	1.78E−15	3.55E−15	−1.78E−15	−7.11E−15	0.00E+00	−3.55E−15	3.55E−15	1.78E−15	−7.11E−15	−1.78E−15	−3.55E−15	0.00E+00	−1.78E−15	5.33E−15	−5.33E−15

Table 1 (continued)

Items and energy terms	Wuhan SARS-CoV-2 Spike/ACE2 (6m0j.pdb)	N501Y	E484K	K417N	K417T	S494P	E484Q	S477N	B.1.427/B.1.429 California (L452R)	B.1.141 N439K	B.1.351 S. Africa (N501Y_E484K_K417N)	B.1.617.1 India (E484Q_L452R)	P.1 Japan/Brazil (N501Y_E484K_K417T)	B.1.1.7_UK (S494P_N501Y_E484K)	B.1.620 (S477N_E484K)
Electrostatic kon (kcal/mol)	−0.18	−0.18	−0.26	0.05	−0.01	−0.17	−0.27	−0.16	−0.32	−0.36	−0.04	−0.20	−0.03	−0.32	−0.29
Energy ionization (kcal/mol)	0.06	0.06	0.07	0.06	0.05	0.00	0.05	0.06	0.05	0.06	0.06	0.05	0.05	0.06	0.06
Entropy complex (kcal/mol)	2.38	2.38	2.38	2.38	2.38	2.38	2.38	2.38	2.38	2.38	2.38	2.38	2.38	2.38	2.38
Number of residues	791	791	791	791	791	791	791	791	791	791	791	791	791	791	791

Table 2 Extract of ACE2 structurally/functionally related proteins sampled by pGenThreader/I-TASSER (see also Supplementary Table 2 for a complete list of the sampled structures). The reported RMSD values between the superimposed/aligned atomic coordi-nates of the compared indicated protein structures were obtained by PyMOL and are reported in Å (1 Å $= 1.0 \times 10^{-10}$ m or 0.1 nm) [63]. Notably, the smaller the RMSD is between two structures, the more similar are the two compared structures [83]

PDB_ID	Chain type	Chains	Target length	Protein name	Virus strain/ infected organism	RMSD (Å) ACE2 (6m0j)	RMSD (Å) ACE1 (6h5w)	RMSD (Å) THOP1 (1s4b)	RMSD (Å) NLN (1i1i)
6m0j	Angiotensin-converting enzyme 2	2	681	Crystal structure of SARS-CoV-2 spike receptor-binding domain bound with ACE2	*Homo sapiens*	0	2.83	4.47	3.9
6vw1	Angiotensin-converting enzyme 2	2	597	Structure of SARS-CoV-2 chimeric receptor-binding domain complexed with its receptor human ACE2	*Homo sapiens*	1.05	2.00	5.03	3.80
1r42	angiotensin I converting enzyme 2	1	615	Native Human Angiotensin Converting Enzyme-Related Carboxypeptidase (ACE2)	*Homo sapiens*	0.48	3.05	5.12	4.33
6h5w	Angiotensin-converting enzyme	1	591	Crystal structure of human Angiotensin-1 converting enzyme C-domain in complex with Omapatrilat	*Homo sapiens*	2.83	0	5.5	5.19
2o3e	Neurolysin	1	678	Crystal structure of engineered neurolysin with thimet oligopeptidase specificity for neurotensin cleavage site	*Rattus Norvegicus*	3.84	5.17	1.09	0.22
1i1iA	Neurolysin	1	681	Neurolysin (endopeptidase 24.16) crystal structure	*Rattus norvegicus*	3.90	5.19	1.12	0
1s4b	Thimet oligopeptidase	1	674	Crystal structure of human thimet oligopeptidase	*Homo sapiens*	4.46	5.50	0	1.12

Table 2 (continued)

PDB_ID	Chain type	Chains	Target length	Protein name	Virus strain/ infected organism	RMSD (Å) ACE2 (6m0j)	RMSD (Å) ACE1 (6h5w)	RMSD (Å) THOP1 (1s4b)	RMSD (Å) NLN (1i1i)
2o36	Thimet oligo-peptidase	1	674	Crystal structure of engineered thimet oligopeptidase with neurolysin specificity in neurotensin cleavage site	*Homo sapiens*	4.05	5.50	0.13	1.05
4fxyP	Neurolysin, mitochondrial	2	693	Crystal structure of rat neurolysin with bound pyrazolidin inhibitor	*Rattus norvegicus*	4.00	4.98	1.24	0.50

close to the "stern" portion of the boat-shaped RBD head (residues T324-L333, ACE2 sequence numbering, and residues E406-L415, THOP1 sequence numbering). A last region with a relative different orientation in the space consists of the THOP1 bent helix (residues M112-K128, THOP1 sequence numbering) corresponding to the ACE2 bent helix (residues L73-Q89, ACE2 sequence numbering) (Fig. 7).

While ACE shows an interaction energy with RBD lower than the ones observed in presence of ACE2, the calculated interaction energies between the SARS-CoV-spike RBD and THOP1 or NLN appear to be stronger than those calculated for ACE2 (Table 3).

Expression of ACE2 structurally related receptors as coronavirus entry factors

GTEx database was screened for estimating the expression levels of ACE2, ACE, THOP1, and NLN together with other SARS-CoV-2 host-cell entry factors in all the tissues available on GTEx. From this analysis, THOP1 resulted more expressed than ACE2 in all the screened tissues (with the exception of kidney cortex, heart left ventricle, adipose visceral, small intestine, showing THOP1/ACE2 similar expression levels) and, most importantly, it is highly expressed in the lung, in the colon, in the esophagus mucosa and in all the brain compartments, in which ACE2 appears poorly expressed (Fig. 8). The expression of NLN or ACE is comparable to, or slightly higher than, ACE2 expression in all the investigated tissues with the exception of heart, adipose tissue, intestine, and testis.

Discussion

The rapid spread of new SARS-CoV-2 variants [1, 6, 48] makes it necessary to develop new tools for evaluating the interactions of SARS-CoV-2 proteins for the host cell receptors. A crucial role in SARS-CoV-2 infection is played by the SARS-CoV-2 spike protein, whose interactions with ACE2 receptor triggers pre-/post-fusion conformational changes causing the virus entry into the human cells [7, 55]. The spike domain responsible for direct interactions with the human ACE2 receptor is the RBD [7, 9, 11, 84–86]. Thus, great attention is dedicated to mutations occurring at the SARS-CoV-2 spike RBD, because they can cause an increase in the binding affinity of the spike protein for the human ACE2 receptor, which may reflect an increased transmissibility or a new acquired antibody escape ability [7, 22–25, 87].

Previously [7], we have determined the interaction energies between the SARS-CoV-2 spike RBD and the human ACE2 receptor available under the PDB_ID 6vw1.pdb [7, 10]. However, the crystallized structure of 6vw1.pdb consisted of a chimeric SARS-CoV-2 spike RBD interacting with ACE2 [10]. The chimeric RBD from 6vw1.pdb showed 22 missense mutations and the deletion of three residues at the RBD with respect to the amino acid sequence of SARS-CoV-2 spike RBD according to the Wuhan reference sequence available under the refseq accession number YP_009724390.1. Less than one year ago, the coordinates of the SARS-CoV-2 spike RBD (according to YP_009724390.1) crystallized in complex with the human ACE2 receptor were made available under the PDB_ID 6mo0j.pdb [88]. Thus, we used the 3D coordinates of 6m0j.pdb for re-estimating the interaction energies at the SARS-CoV-2 spike RBD/ACE2 protein–protein interface and we modelled RBD mutants on the 6m0j.pdb protein template and quantified the affinity of the known RBD mutants for the human ACE2 receptor. More in detail we studied the impact of the mutations responsible for the investigated VoC

Fig. 5 Superimposition of the ACE2 structurally related proteins ACE, THOP1, and NLN to ACE2. The superimposition of the three ACE2 structurally related proteins to ACE2 in complex with the SARS-CoV-2 spike RBD allows to highlight the overall structural similarity between ACE, ACE2, NLN, and THOP1 and based on the observed structural similarity it is possible to speculate about a possible interaction between the three ACE2 structurally related proteins and the SARS-CoV-2 spike RBD, that might be favored by future spike variants showing mutations at the SARS-CoV-2 spike RBD. **A**–**D** four inspections of the lateral view rotated of 90° (along the RBD z axis) within each column showing the superimposition of ACE (6h5w.pdb, magenta cartoon) on ACE2 ((6m0j.pdb, yellow cartoon) interacting with the SARS-CoV-2 spike RBD (6m0j.pdb, light grey cartoon). **E**–**H** report the corresponding THOP1/ACE2 superimposition (THOP1, 1s4b.pdb, green cartoon), whereas **I**–**L** report the corresponding NLN/ACE2 superimposition (NLN, 1i1i.pdb, cyan cartoon)
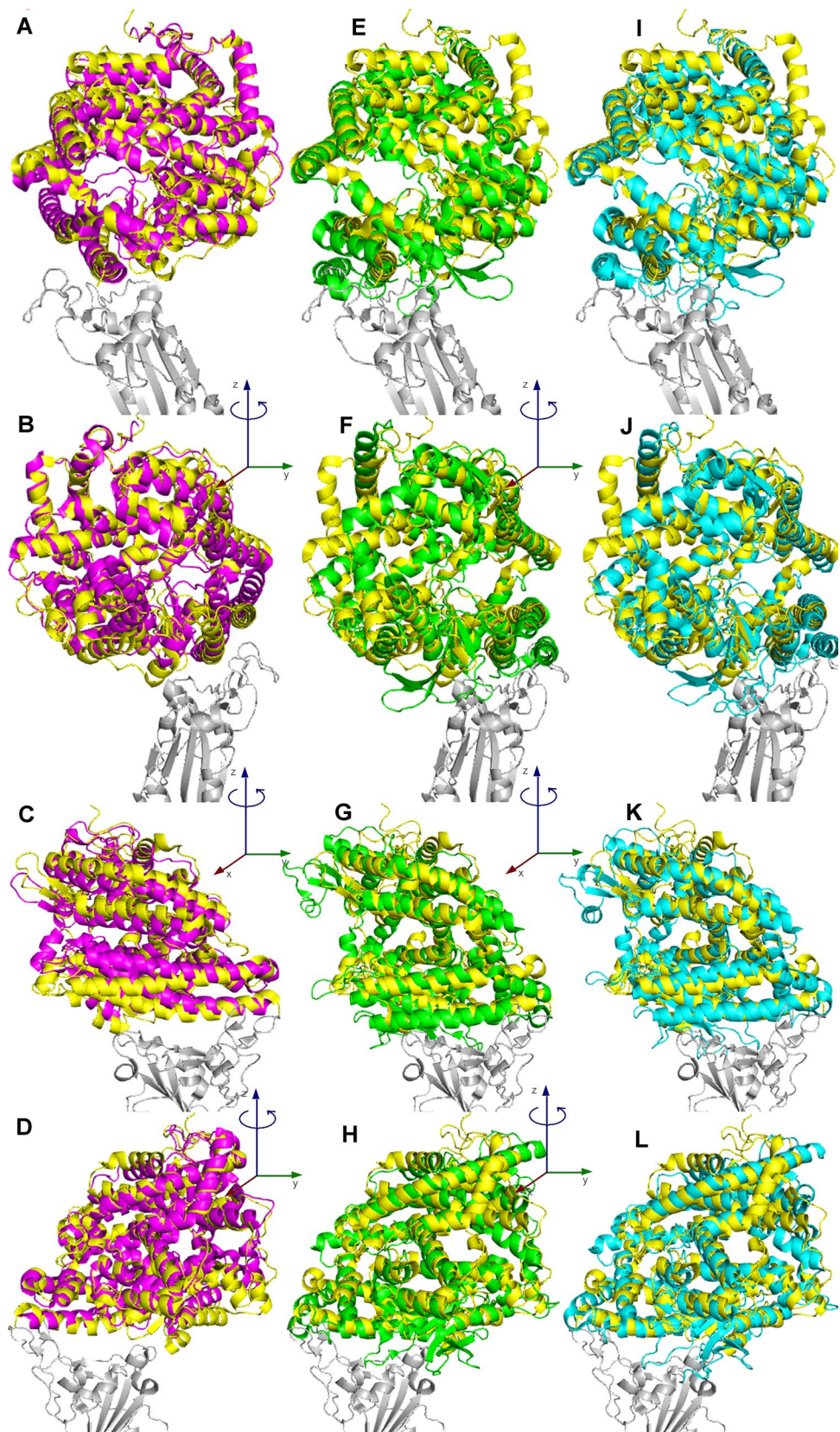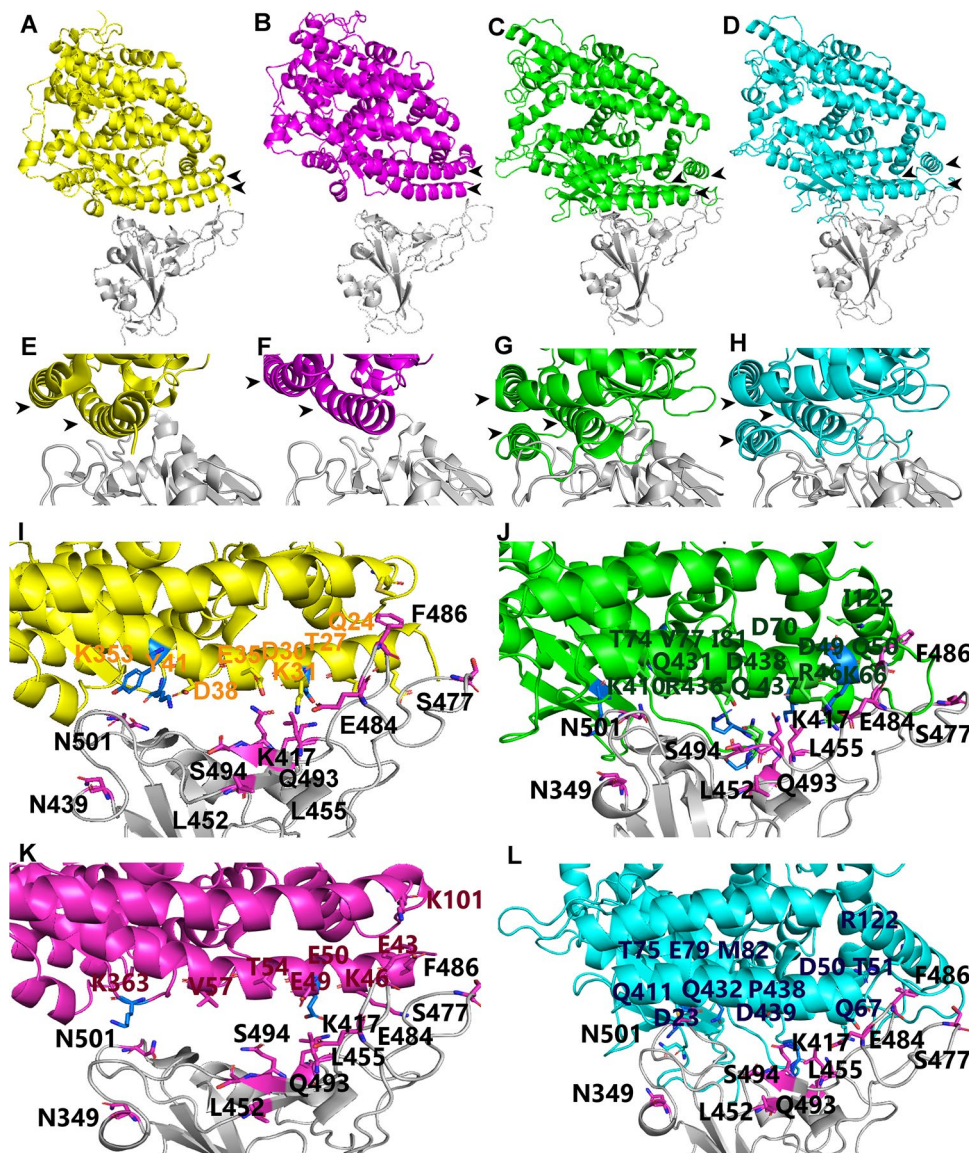
Fig. 6 Comparative structural analysis of the ACE2 structurally related proteins ACE, THOP1, and NLN. The superimposition of the three ACE2 structurally related proteins to ACE2 in complex with the SARS-CoV-2 spike RBD has also allowed to highlight the putative secondary structure elements and specific residues involved in possible interactions between the SARS-CoV-2 spike RBD and the three ACE2 structurally related proteins, ACE, NLN and THOP1. The SARS-CoV-2 spike RBD (light gray cartoon) complexed with **A** ACE2 (yellow cartoon), **B** ACE (magenta cartoon), **C** THOP1 (green cartoon), and **D** NLN (cyan cartoon). **E–H** Zoomed views of the SARS-CoV-2 spike RBD interacting with ACE2, ACE, THOP1, and NLN. Black arrowheads in **A–H** indicate secondary structure elements (alpha-helix) mainly involved in interactions with the SARS-CoV-2 spike RBD. **I–L** Zoomed views of the RBD (light grey cartoon) showing the RBD amino acids K417; N439; L452; E484; S494; N501 (magenta sticks) observed mutated in the investigated VoC. **I** Residues within 4 Å from SARS-CoV-2 spike RBD highlighted within ACE2 (yellow), **K** ACE (magenta), **J** THOP1 (green), and **L** NLN (cyan) are reported in colored sticks. ACE2 structurally related receptor residues within 4 Å from the SARS-CoV-2 spike RBD protein positions observed mutated in the investigated VoC are reported in blue sticks (see Supplementary Table 1 for a list of the detected interactions between the Wuhan SARS-CoV-2 spike RBD and ACE2 structurally related receptors according to PIC estimations)

B.1.1.7-UK (carrying the mutations of concern/interest N501Y, S494P, E484K), P.1-Japan/Brazil (K417N/T, E484K, N501Y), B.1.351-S. Africa (K417N, E484K, N501Y), B.1.427/B.1.429-California (L452R), the B.1.141 (N439K) and the recent B.1. 617-India (L452R, E484Q) VoC. The effect of the investigated mutations on the SARS-CoV-2 spike RBD structure and on RBD/ACE2 interactions was checked revealing that all of them might perturb the RBD structure either at the level of the secondary structure elements hosting the investigated mutations or at the level of spatially close secondary structure elements on the entire RBD head.

Fig. 7 A, **B** Zoomed views of the SARS-CoV-2 spike RBD head (white cartoon) in contact with the ACE2 structurally related proteins. The superimposed ACE2, ACE, THOP1 and NLN are reported in yellow, magenta, green and cyan cartoon, respectively. **C–H** Zoomed views of secondary structural elements(alpha helices or beta sheets) highlighted from the analyzed ACE2 structurally related proteins. Residues ad the beginning and at the end of the reported secondary structural elements are indicated by labels. MSA of the amino acid sequences of the corresponding highlighted secondary structural elements are reported below each highlighted structural element of **C–H**. The zoomed views of the superimposition of the three ACE2 structurally related proteins to ACE2 in complex with the SARS-CoV-2 spike RBD allow to highlight the structural similarity between the reported secondary structure elements, despite of the variability of the amino acid composition of the same structure elements, as observed in the corresponding sequence alignment panels

Table 3 FoldX interaction energies between SARS-CoV-2 RBD and ACE2 structurally related proteins. Please note that more negative "interaction energy" values, reported here in kcal/mol, indicate stronger interaction energies and thus higher binding affinity at the protein–protein interface of the indicated SARS-CoV-2 spike RBD/ACE2 or spike RBD/ACE2 structurally related protein complexes [7, 73]. The energy terms and contributions are reported in kcal/mol according to FOLDX indications. Each energy term has a specific weight in the calculation of the interaction energy, according to the equations reported in [73]. The calculated energy terms and item names are reported in the first cell of each row (i.e., in the first column of the reported table), whereas items (protein names, chain-ID, or protein PDB accession codes) and the calculated energy terms are reported for ACE2 and the indicated ACE2 structurally related proteins within columns 2–5. The energy terms and contributions are reported in kcal/mol according to FOLDX indications For a complete explanation of item names and energy terms, please, visit the link http://foldxsuite.crg.eu/command/AnalyseComplex. Interaction energies are also reported in kJ/mol (1 kcal = 4.184 kJ)

Items and energy terms	ACE2 (6m0j.pdb)	ACE (6h5w.pdb)	NLN (1i1i.pdb)	THOP1 (1s4b.pdb)
Group 1 (chain)	A (ACE2)	A (ACE1)	P (NLN)	P (THOP)
Group 2 (chain)	E (RBD_6m0j)	E (RBD_6m0j)	E (RBD_6m0j)	E (RBD_6m0j)
Interaction energy (kJ/mol)	**− 85.85**	**− 31.02**	**− 99.29**	**− 92.09**
Interaction energy (kcal/mol)	**− 20.51**	**− 7.41**	**− 23.72**	**− 22.00**
Intraclashes group 1	22.28	22.95	21.50	22.95
Intraclashes group 2	5.49	6.63	13.07	9.70
Backbone H-bond (kcal/mol)	− 3.64	− 0.16	− 13.08	− 7.86
Sidechain H-bond (kcal/mol)	− 12.12	− 9.76	− 19.3	− 9.39
Van der Waals (kcal/mol)	− 15.89	− 7.81	− 55.11	− 48.08
Electrostatics (kcal/mol)	− 2.41	− 2.65	− 1.62	− 3.31
Solvation polar (kcal/mol)	22.14	12.62	84.22	76.28
Solvation hydrophobic (kcal/mol)	− 20.02	− 9.90	− 68.74	− 60.04
Van der Waals clashes (kcal/mol)	0.41	0.06	6.96	2.91
Entropy side chain (kcal/mol)	9.53	9.02	18.80	18.02
Entropy main chain (kcal/mol)	1.57	1.33	21.16	8.86
Torsional clash (kcal/mol)	0.083	0.055	3.68	0.94
Backbone clash (kcal/mol)	2.45	0.79	17.26	14.11
Helix dipole (kcal/mol)	− 0.06	− 0.02	− 0.00	0.14
Disulfide (kcal/mol)	1.78E − 15	3.55E − 15	0	0
Electrostatic kon (kcal/mol)	− 0.18	− 0.20	− 0.67	− 0.47
Energy Ionization (kcal/mol)	0.07	− 2.84E − 16	7.06E − 03	1.60E − 16
Entropy complex (kcal/mol)	2.38	2.38	2.38	2.38
Number of residues	791	780	859	848

Fig. 8 Expression of SARS-CoV-2 host-cell entry factors. Heatmap analysis about the expression of several SARS-CoV-2 host-cell entry factors and ACE2 structurally related proteins was obtained through GTEx multi-gene query expression (GTEx based on V8 Release, for more details, please visit https://gtexportal.org/home/). The "*" symbols indicate the position of the main discussed tissues (brain, esophagus mucosa, lung). The "°, #, v, ^, $" symbols indicate THOP1, ACE2, Furin, ACE, NLN expression levels, respectively. TPM: transcripts per million

Evaluation of the interaction energies at the spike RBD/ACE2 interface through the proposed modular molecular framework

Through the monitoring of the binding interactions, it is possible to see that H-bonds and hydrophobic interactions among side chains at the protein–protein interface, but also backbone–side chain H-bonds, among the very short-range (< 3.8 Å) interactions [42], can substantially increase or decrease as a consequence of a mutation occurring at the boat-shaped RBM of the SARS-CoV-2 spike RBD, as observed for the single N439K amino acid replacement occurring at the "stern" RBM region, or for the B.1.141 VoC or in the P.1 Japan/Brazil VoC consisting of the triple mutant N501Y_E484K_K417T showing mutations along the entire RBM, with respect to the interactions calculated in the crystallized SARS-CoV-2 spike RBD/ACE2 protein complex (6m0j.pdb) used as a reference structure. The replacement of N501 with a tyrosine introduces new aromatic-aromatic and π-aromatic binding interactions in the short/medium-range (3.8–9.5 Å) interactions [42], without apparently increasing SARS-CoV-2 spike RBD/ACE2 binding affinity.

By monitoring binding affinity and interaction energies, it appears that the single amino acid replacements N439K (B.1.141 VoC) and E484K (detected in several VoC) cause the most dramatic increase in interaction energies among the investigated single mutants. In addition, the triple mutant N501Y_E484K_K417T detected in the P.1 Japan/Brazil VoC [24, 52] shows an increase of about 5% in binding affinity that might reflect the greater ability shown by this VoC in escaping antibodies produced as a consequence of SARS-CoV-2 Wuhan-sequence based vaccination [24]. Also, the B.1.1.7_UK S494P_N501Y_E484K investigated variant shows an important variation in the interactions at the ACE2/RBD interface and a decrease in the calculated interaction energies. Notably, this VoC shows the replacement of S494 with a proline residue. P, S, and G play a hinge role in local secondary structures especially when other P, S or G are near in the sequence [78, 89]. The replacement of a serine with a proline may confer less flexibility to the local secondary structure and thus it is retained this mutation limits RBD local flexibility causing a decrease in the number of short/medium-range interactions [42] at the ACE2/RBD interface. It can be argued that the limited flexibility introduced by the S494P amino acid replacement decreases the affinity for ACE2 and makes sensitive this VoC to antibodies produced by vaccination [52], despite of the presence of N501Y and E484K that as single mutants show a higher affinity for the ACE2 receptor. A similar reduced flexibility might be expected for the RBD of the double mutant S477N_E484K due to the replacement of the Ser 477 with an Asn residue. Our conclusions about changes in interaction energies and number of interactions for the investigated VoC

are coherent with the expected effects about spike RBD/ACE2 interactions and antibody escape reported for the described VoC [22–25, 49–52]. In order to better link binding affinity, transmissibility, virulence and disease severity, it would be useful to associate the calculated SARS-CoV-2 spike/ACE2 interaction energies with both the number of infected people with a specific variant and the disease severity (when possible).

Controversial *in vitro* binding assays for gaining clues about VoC/VoI virulence and transmissibility

While several studies about in vitro ACE2/spike binding assays are becoming available, an in vitro golden standard technique for quickly estimating the interaction energies between ACE2 and RBD variants is difficult to establish due to relative differences observed between the binding affinity estimated for the ACE2/Wuhan spike protein–protein complex and those estimated for ACE2/RBD variants ([90–92] and Supplementary Table 3). The observed differences can be ascribed to the employment of the only RBD instead of the entire spike protein, as well as to the employment of living cells expressing ACE2 (full length) or recombinant truncated human ACE2 ectodomain. Also, the different employed ACE2/RBD ratio in the performed binding assays, as well as the different probes and employed detectors can be related to the observed differences in binding affinity (Supplementary Table 3).

Interpretation of the results of our in silico approach for the prediction of VoC/VoI transmissibility and virulence, based on the predicted SARS-CoV-2 spike RBD/ACE2 binding affinity

Among the investigated VoC-RBDs, the triple RBD mutant of the P.1 Japan/Brazil VoC shows the highest increase in the binding affinity (a 5% increase) compared to the binding affinity estimated for the Wuhan spike RBD/ACE2 protein complex. All the other mutants show lower increases or decreases in the calculated interaction energies, compared to the Wuhan spike RBD/ACE2 protein complex. These considerations make us hypothesize that an increase of less than 5% in the calculated interaction energies between the VoC-RBDs and ACE2, compared to interactions energies calculated for the Wuhan spike RBD/ACE2 protein complex, will not dramatically change the virulence of new SARS-CoV-2 VoC, making them still sensitive to the current employed Wuhan SARS-CoV-2 based vaccines [29, 93], although a slightly increased VoC-RBD/ACE2 binding affinity might determine a lower affinity for the vaccine induced antibodies [22–24]. Conversely, VoC showing acidic/hydrophilic amino acid replacement with basic residues at the "stern" and "bow" regions of the boat-shaped RBM on the spike RBD, or more than 3 simultaneous mutations along the entire RBM on the spike RBD head, and/or

an increase of more than 5–10% in the calculated interaction energies with ACE2 should deserve great attention.

In this context interaction energies quickly calculated by using the described molecular framework can be used for leading/driving in vitro binding assays, as a more replicable in vitro binding assay will be established, for estimating the putative aggressiveness of new VoC, and for the development of new vaccines and antibodies.

Furthermore, while trying to understand if there was an underestimation of SARS-CoV-2 preexisting immunity [94], in the context of personalized medicine, it should be stressed that our molecular framework can be used for investigating and explaining the resistance shown by some individuals to the infection or to the development of clinical manifestations despite infection, by evaluating the importance of single gene variants in the cell entry factors [59] shown by resistant individuals, starting from putative ACE2 variants [39, 95]. At the same time this kind of analysis might help in drawing new recombinant ACE2 proteins with very high affinity for spike RBD variants to be administered to patients for blocking or slowing down the infection [96].

Fold recognition tools for predicting ACE2 structurally related alternative cell-entry factors

In the end, we wondered about the possible existence of ACE2 structurally related receptors that might interact with SARS-CoV-2 spike RBD in those tissues showing ACE2 poorly expressed [59, 60]. Our fold-recognition based analysis revealed that beyond the expected ACE, THOP1, and NLN appear structurally related to ACE2. ACE, THOP1, and NLN share the "oligopeptidase/receptor" activity with ACE2 and show at least one isoform localized at the plasma membrane [97, 98]. While ACE appears to form a weaker protein complex with the SARS-CoV-2 spike RBD, in our interaction analyses, the interaction energies calculated at the protein–protein interface of the SARS-CoV-2 spike RBD/THOP1 or SARS-CoV-2 spike RBD/NLN show binding energies higher than those calculated for the SARS-CoV-2 spike RBD/ACE2 protein complex, maybe due to a supplementary alpha-helix located at the interface with the SARS-CoV-2 spike RBD in the obtained 3D protein complexes. NLN and THOP1 participate to the cleavage of cytosolic peptides and share the 80% of identical residues [98].

While the expression of NLN or ACE is comparable to or slightly higher than ACE2 expression in all the investigated tissues, it is possible to see that THOP1 is highly more expressed than ACE2 in most of the screened tissues and, most importantly, it is highly expressed in the lung, in the colon, in the esophagus mucosa and in all the brain compartments.

Although THOP1 and ACE2 show a low percentage of identical residues (< 30%), their related biochemical function

and high structural similarity (overall RMSD < 4.5 Å) make THOP1 a suggestive/alternative candidate receptor for the SARS-CoV-2 spike protein. In addition, in our GTEx analyses the high expression of THOP1 appears to correlate with high expression levels of the Furin protease, which may participate to spike cleavage allowing pre-/post-fusion conformational changes crucial for host-cell penetration [59, 99], above all in those tissues showing a low ACE2 expression, i.e., in the lung, in the esophagus mucosa or in the brain. Notably, THOP1 expression was observed upregulated in COVID-19 infected patients and in particular in the effector CD8 T-cells of COVID-19 patients at the beginning of the infection [100]. Furthermore, THOP1 appears to play a crucial role in the regulation of MHC I cell-surface expression [101–103]. Due to the shown high expression of THOP1 in several brain compartments and in the effector CD8 T-cells of COVID-19 patients [100], it raises the question about a possible relationship between different SARS-CoV-2 immune responses [100, 104], neurological disorders observed in long covid patients, and THOP1 involvement in MHC-class I regulation in COVID-19 patients [105–107].

Employment of the presented pipeline in the routine hospital laboratories

While waiting for a definitive quick in vitro binding assay for estimating SARS-CoV-2 variant transmissibility, virulence, and disease severity, the presented approach may represent a valid tool for gaining clues about new virus strain aggressiveness. The presented strategy just consists of a computational pipeline with no consumable costs, needing only a good internet connection for using the webservice, as long as the hospital laboratory is able to provide genome sequencing data (in the context of personalized medicine). The employment of SARS-CoV-2 genome sequencing in combination with a pipeline like the ones here presented would help in quickly highlighting new SARS-CoV-2 spike variants with high affinity for ACE2 to be carefully monitored and that is the reason for which similar pipelines should be routinely used in clinics in this pandemic era. Indeed, the early recognition of a SARS-CoV-2 spike variant with high affinity for the human ACE2 would alert clinicians, who would then be able to deal with patients in the safest way or to contact people positive to the test to provide appropriate support and indications, also in case of asymptomatic/paucisymptomatic people carrying dangerous variants. On this concern, it would also be useful to sequence ACE2 in all those patients/carriers of new SARS-CoV-2 spike variants, which would help in calculating the binding affinity at the SARS-CoV-2 spike variant/ACE2 variant protein–protein interface in a more detailed way, by using our molecular framework. This would also help in explaining a variation in binding affinity as well as in collecting data about a putative innate (partial

or complete) immunity, depending on specific mutations also on ACE2 and/or on the severity of disease manifestations. The employment of our molecular framework in combination with the ongoing SARS-CoV-2 genome sequencing activities, as proposed, would help in this way clinicians to put in play all the necessary precautionary counteractions for preventing (or at least slowing down) the spread of predicted high virulent variants.

Strengths and limitations

In this study, the binding interactions between the spike RBD and ACE2 were systematically analyzed through the presented modular molecular framework for attempting to create a transmissibility, virulence and disease severity prediction tool based on the variation of the calculated spike RBD/ACE2 binding affinities, whose increase indicates more stable interactions between the SARS-CoV-2 spike protein and the human ACE2 receptor, which might reflect an increased ability of specific variants in penetrating host-cells. We also made available for clinicians and researchers a webservice (https://www.mitoairm.it/covid19affiniti es), which they can easily use by uploading specific SARS-CoV-2 spike RBD (or the entire spike) sequences and/or ACE2 sequences for quickly calculating the interaction energies at the SARS-CoV-2 spike RBD/ACE2 protein–protein interface. The interaction energies calculation provides a binding affinity estimation which might be correlated to the VoC virulence/transmissibility based on the comparison with the binding affinities and interaction energies calculated at the spike RBD/ACE2 protein–protein interface for the VoC analyzed in this manuscript. The strength of our pipeline is that it is easy to use through our webservice and gives results about the investigated interaction energies in 24–48 h for each single analysis.

One limitation is represented by the current computational power (a high number of requests would slow-down results receipt) given that the calculations are mainly performed on a small local server.

Another limitation is now represented from the fact that our pipeline was trained on mutations occurring at the SARS-CoV-2 spike RBD. Indeed, while it is expected that mutations at the RBD need to be monitored because those mutations can be responsible for an increase in the binding affinity for the ACE2 receptor, it is matter of debate the increased efficiency in entering host-cells proposed for variants showing supplementary mutations far from the RBD, i.e., D614G or P681H shown by several VoC [36] and other variants of interest or under monitoring (https://www.ecdc. europa.eu/en/covid-19/variants-concern). It is retained that mutations like D614G and P681H, being located in the N-terminal portion of the spike pre-fusion conformation,

despite of their relatively great distance from the SARS-CoV-2 spike RBD/ACE2 protein–protein interface, can confer a different flexibility (increased for the D614G mutant and decreased for the P681H mutant) to the entire spike protein, before cleavage events determining the post-fusion conformation, following the spike N-terminal loss [7]. Indeed, it was observed that a successful interaction of the spike protein with the ACE2 receptor is supported by the high flexibility of the SARS-CoV-2 spike protein in the pre-fusion conformation, while scanning host-cell surface [8, 9].

On this concern, a point of strength of our approach is that it can be easily adapted for estimating ACE2/spike interactions by using the entire spike protein instead of the only RBD, for investigating the effect of spike mutations occurring far from the RBD on interactions with ACE2. Furthermore, the analysis of the interactions between the entire spike protein and ACE2 would allow to evaluate the binding affinity in various combinations because it is known that the number of interacting ACE2/spike proteins might vary depending on the number of RBD in up-/down-conformation [108] (i.e., the theoretical maximum number of ACE2/spike interacting protein elements consists of 2 spike (trimeric) proteins interacting with 6 ACE2 proteins, as previously predicted [7] and then observed in the cryo-EM structure 7kni. pdb [108]). Once obtained this multi-subunit complexes, the entire protein complexes can be relaxed by using molecular dynamics [109, 110] for the following estimation of the interaction energies at the ACE2/spike protein interface. The possibility to relax and study the entire trimeric spike protein interacting with ACE2 (and/or antibodies) [7] might allow to predict more efficiently the aggressiveness of future VoC and/or their abilities in escaping vaccine induced antibodies and this would help in obtaining better predictions in the context of personalized medicine [15, 16, 111–113]. In addition, results of the analyses performed on the entire spike protein interacting with ACE2 and/or antibodies would also allow to design more efficient antibodies which will target specific portions of the spike RBD for preventing ACE2 binding, without creating at the same time undesired clashes in protein structure with different chains of the same spike trimeric protein. However, in order to perform the presented in silico binding assays by using the entire spike protein and/or the trimeric spike would surely need more computational power.

Yet, another limitation of our approach is currently represented by the need to train our simulator with a very high number of SARS-CoV-2 spike and/or ACE2 variant combinations to be analyzed and sorted by binding energies to be associated with the disease severity observed in patients who undergo SARS-CoV-2 genome (and ACE2) sequencing analysis. The association of binding energies with disease severity would help in making more accurate transmissibility, virulence and disease severity predictions based not only on

the calculated interaction energies at the SARS-CoV-2 spike RBD/ACE2 protein–protein interface, but also on the localization of the mutation and on the specific interactions that a spike variant can establish with a specific ACE2 variant.

A last point of strength of our computational approach is based on the ability of our pipeline to detect ACE2 structurally related proteins to be monitored as alternative cell-entry factors in those cells showing a low expression of ACE2. The highlighted structural relationships between ACE2, ACE, THOP1, and NLN might be useful for the scientific community that should monitor and evaluate the putative abilities of future SARS-CoV-2 variants and other related coronaviruses to use ACE2 structurally related proteins to penetrate host-cells, which might result even more dangerous due to the role of THOP1 in the regulation of MHC I cell-surface expression [105–107].

Conclusions and recommendations

In this study, we have presented and tested a modular molecular framework for quickly estimating the binding affinity of the human ACE2 reference protein (NP_001358344.1) for the SARS-CoV-2 spike RBD protein variants, as sequenced from 7 VoC. Although, several single or multiple mutations at the spike RBD, as detected in the investigated VoC, did not produce an increase in the strength of interaction energies at the spike RBD/ACE2 protein–protein interface, it was verified that particular single mutations at specific regions of the "boat-shaped" RBM, on the spike RBD, can increase the interaction energies at the protein–protein interface with ACE2. The observed higher interaction energy can implicate a greater binding affinity, which might reflect an increased ability for the analyzed VoC to penetrate host-cells.

In order to gain information about the binding affinity of new spike RBD variants for the ACE2 receptor, we would suggest using the webservice that we made available for clinicians and researchers, who can upload their sequenced RBD variants (and ACE2 variants, if possible), to calculate the interaction energies at the RBD/ACE2 protein–protein interface among sequences obtained along SARS-CoV-2 genome sequencing activities in patients and/or people who requested for molecular diagnosis tests.

In addition, with the right computational resources, the presented pipeline can be employed for building a library of all the possible RBD variants, which might be predicted based on solid evolutionary tools [114–116] and population genetics tools [117, 118], for trying to understand which (and where) future spike variants will appear [2, 5, 16], to quickly predict their binding affinity for the different known ACE2 variants present in the world population (i.e., see https://gnomad.broadinstitute.org/gene/ENSG00000130234?dataset=exac). Thus, it would be possible to

early predict which RBD variants might be more dangerous for human health even before their appearance. This would allow to prepare all the possible preventive actions in case of appearance of the worst variants, preventing the most deleterious scenarios of new outbreaks.

Furthermore, we would recommend using our webservice for the estimation of the interaction energies at the spike RBD/ACE2 protein–protein interface, for each RBD/ACE2 combination, as sequenced from patients, would allow creating a list of the RBD/ACE2 protein complexes sorted by increasing interaction energy and the corresponding degree of disease severity for each known variant combination. The interaction energy parameter, function of the disease severity for each patient, synergically with other established priority parameters [111, 119], might be used for training our pipeline to provide more and more accurate predictions about future variants transmissibility, virulence and disease severity. The early recognition of more transmissible/virulent variants would also help in assigning priority and personalized treatments to future patients, based on the specific RBD/ACE2 variants sequenced from the infected patients, by overcoming most of the contradictory results of the currently employed in vitro assays. This would also help in planning more personalized clinical treatments for those patients showing a known combination of spike/ACE2 variants, already successfully treated from other clinicians [15]. Considering this, it would be useful in the future to build a free and publicly accessible database of all the screened spike/ACE2 variant combinations (as detected from patients) with the corresponding binding affinity parameters, associated to disease severity, clinical treatments and outcome, for helping clinicians in treating patients, in accordance with the vision and attitudes of PPPM/3PM.

Data availability All data generated or analyzed during this study are included in this published article [and its supplementary information files].

Code availability All the necessary information are provided along the manuscript.

Declarations

Ethics approval NA

Consent to participate NA

Consent for publication NA

Conflict of interest The authors declare no competing interests.

References

1. Yuan F, Wang L, Fang Y, Wang L. Global SNP analysis of 11,183 SARS-CoV-2 strains reveals high genetic diversity. Transbound Emerg Dis. 2021;68(6):3288–304. https://doi.org/10.1111/tbed.13931.
2. Kupferschmidt Kai. New SARS-CoV-2 variants have changed the pandemic. What will the virus do next? Science. 2021;373:6557.
3. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;579:270–3.
4. Fang S, Li K, Shen J, Liu S, Liu J, Yang L, et al. GESS: a database of global evaluation of SARS-CoV-2/hCoV-19 sequences. Nucleic Acids Res. 2021;49:D706–14.
5. Pucci F, Rooman M. Prediction and evolution of the molecular fitness of sars-cov-2 variants: Introducing SpikePro. Viruses. 2021;13:935.
6. Chiara M, D'Erchia AM, Gissi C, Manzari C, Parisi A, Resta N, et al. Next generation sequencing of SARS-CoV-2 genomes: challenges, applications and opportunities. Brief Bioinform. 2021;22:616–30.
7. Mercurio I, Tragni V, Busto F, De Grassi A, Pierri CL. Protein structure analysis of the interactions between SARS-CoV-2 spike protein and the human ACE2 receptor: from conformational changes to novel neutralizing antibodies. Cell Mol Life Sci. 2021;78:1501–22.
8. Pierri CL. SARS-CoV-2 spike protein: flexibility as a new target for fighting infection. Signal Transduct Target Ther. 2020;5:4–6.
9. Turoňová B, Sikora M, Schürmann C, Hagen WJH, Welsch S, Blanc FEC, et al. In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. Science. 2020;370:203–8.
10. Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, et al. Structural basis of receptor recognition by SARS-CoV-2. Nature. 2020;581:221–4.
11. Wrapp D, Wang N, Corbett KSS, Goldsmith JAA, Hsieh C-L, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science. 2020;367:1260–3.
12. Zhang L, Lin D, Sun X, Curth U, Drosten C, Sauerhering L, et al. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. Science. 2020;368:409–12.
13. Elfiky AAA. Ribavirin, Remdesivir, Sofosbuvir, Galidesivir, and Tenofovir against SARS-CoV-2 RNA dependent RNA polymerase (RdRp): A molecular docking study. Life Sci. 2020;253:11759.
14. Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, O'Meara MJ, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature. 2020;583:459–68.
15. Wang LYY, Cui JJJ, OuYang QYY, Zhan Y, Wang YMMY, Xu XYY, et al. Complex analysis of the personalized pharmacotherapy in the management of COVID-19 patients and suggestions for applications of predictive, preventive, and personalized medicine attitude. EPMA J. 2021;12:307–24.
16. Zahradník J, Marciano S, Shemesh M, Zoler E, Harari D, Chiaravalli J, et al. SARS-CoV-2 variant prediction and antiviral drug design are enabled by RBD in vitro evolution. Nat Microbiol. 2021;6:1188–98.
17. Lurie N, Saville M, Hatchett R, Halton J. Developing Covid-19 Vaccines at Pandemic Speed. N Engl J Med. 2020;382:1969–73.
18. Chen WH, Strych U, Hotez PJ, Bottazzi ME. The SARS-CoV-2 Vaccine Pipeline: an Overview. Curr Trop Med Rep. 2020;3:1–4. https://doi.org/10.1007/s40475-020-00201-6.
19. Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. N Engl J Med. 2020;383:2603–15.
20. Baden LR, El Sahly HM, Essink B, Kotloff K, Frey S, Novak R, et al. Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. N Engl J Med. 2021;384:403–16.
21. Vandeputte J, Van Damme P, Neyts J, Audonnet JC, Baay M, Neels P. Animal experiments show impact of vaccination on reduction of SARS-CoV-2 virus circulation: a model for vaccine development? Biologicals. 2021;73:1–7.
22. Thomson EC, Rosen LE, Shepherd JG, Spreafico R, da Silva FA, Wojcechowskyj JA, et al. Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. Cell. 2021;184:1171-1187.e20.
23. Greaney AJ, Loes AN, Crawford KHD, Starr TN, Malone KD, Chu HY, et al. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. Cell Host Microbe. 2021;29:463-476.e6.
24. Zhou D, Dejnirattisai W, Supasa P, Liu C, Mentzer AJ, Ginn HM, et al. Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera. Cell. 2021;184:2348-2361.e6.
25. McCallum M, Bassi J, De Marco A, Chen A, Walls AC, Di Iulio J, et al. SARS-CoV-2 immune evasion by variant B.1.427/B.1.429. Science. 2021;373:648–54.
26. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob Challenges. 2017;1:33–46.
27. Leung NHL. Transmissibility and transmission of respiratory viruses. Nat Rev Microbiol. 2021;19:528–45.
28. Geoghegan JL, Holmes EC. The phylogenomics of evolving virus virulence. Nat Rev Genet. 2018;19:756–69.
29. Kustin T, Harel N, Finkel U, Perchik S, Harari S, Tahor M, et al. Evidence for increased breakthrough rates of SARS-CoV-2 variants of concern in BNT162b2-mRNA-vaccinated individuals. Nat Med. 2021;27:1379–84.
30. Khoury DS, Wheatley AK, Ramuta MD, Reynaldi A, Cromer D, Subbarao K, et al. Measuring immunity to SARS-CoV-2 infection: comparing assays and animal models. Nat Rev Immunol. 2020;20:727–38.
31. Srivastava VK, Kaushik S, Bhargava G, Jain A, Saxena J, Jyoti A. A Bioinformatics approach for the prediction of immunogenic properties and structure of the SARS-COV-2 B.1.617.1 variant spike protein. Biomed Res Int. 2021;2021:7251119. https://doi.org/10.1155/2021/7251119.

32. Periwal N, Rathod SBB, Pal R, Sharma P, Nebhnani L, Barnwal RPP, et al. In silico characterization of mutations circulating in SARS-CoV-2 structural proteins. J Biomol Struct Dyn. 2021;1–16. https://doi.org/10.1080/07391102.2021.1908170.

33. Yang HM, Junior LPL, Yang AC. Evaluating the trade-off between transmissibility and virulence of SARS-CoV-2 by mathematical modeling. medRxiv. 2021. https://doi.org/10.1101/2021.02.27.21252592.

34. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell Elsevier. 2020;182:812-827.e19.

35. Alizon S, Sofonea MT. SARS-CoV-2 virulence evolution: avirulence theory, immunity and trade-offs. J Evol Biol. 2021;00:jeb.13896.

36. Ozono S, Zhang Y, Ode H, Sano K, Tan TSS, Imai K, et al. SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity. Nat Commun. 2021;12:848.

37. Bhattacharjee MJ, Lin JJ, Chang CY, Chiou YT, Li TN, Tai CW, et al. Identifying primate ACE2 variants that confer resistance to SARS-CoV-2. Mol Biol Evol. 2021;38:2715–31.

38. Benetti E, Tita R, Spiga O, Ciolfi A, Birolo G, Bruselles A, et al. ACE2 gene variants may underlie interindividual variability and susceptibility to COVID-19 in the Italian population. Eur J Hum Genet. 2020;28:1602–14.

39. Casanova JL, Su HC, Abel L, Aiuti A, Almuhsen S, Arias AA, et al. A global effort to define the human genetics of protective immunity to SARS-CoV-2 infection. Cell. 2020;181:1194–9.

40. Wan Y, Shang J, Graham R, Baric RSS, Li F. Receptor Recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. J Virol. 2020;9:e00127-e220.

41. Tanaka S, Scheraga HA. Model of protein folding: inclusion of short-, medium-, and long-range interactions. Proc Natl Acad Sci U S A. 1975;72:3802–6.

42. Onofrio A, Parisi G, Punzi G, Todisco S, Di Noia MAA, Bossis F, et al. Distance-dependent hydrophobic-hydrophobic contacts in protein folding simulations. Phys Chem Chem Phys. 2014;16:18907–17.

43. Koliński A, Bujnicki JM. Generalized protein structure prediction based on combination of fold-recognition with de novo folding and evaluation of models. Proteins. 2005;61:84–90.

44. Karplus M, Sali A. Theoretical studies of protein folding and unfolding. Curr Opin Struct Biol. 1995;5:58–73.

45. Reinhard A, Nürnberger T. Steady-state and kinetics-based affinity determination in effector-effector target interactions. Methods Mol Biol. 2017;1578:81–108.

46. Brito AF, Pinney JW. Protein-protein interactions in virus-host systems. Front Microbiol. 2017;8:1–11.

47. Wallqvist A, Memišević V, Zavaljevski N, Pieper R, Rajagopala SV, Kwon K, et al. Using host-pathogen protein interactions to identify and characterize Francisella tularensis virulence factors. BMC Genomics. 2015;16:1–18.

48. Chiara M, Zambelli F, Tangaro MA, Mandreoli P, Horner DS, Pesole G. CorGAT: a tool for the functional annotation of SARS-CoV-2 genomes. Bioinformatics. 2021;36:5522–3.

49. Yadav PD, Sapkal GN, Abraham P, Ella R, Deshpande G, Patil DY, et al. Neutralization of Variant Under Investigation B.1.617.1 With Sera of BBV152 Vaccinees. Clin Infect Dis. 2021;ciab411. https://doi.org/10.1093/cid/ciab411.

50. Frampton D, Rampling T, Cross A, Bailey H, Heaney J, Byott M, et al. Genomic characteristics and clinical effect of the emergent SARS-CoV-2 B.1.1.7 lineage in London, UK: a whole-genome sequencing and hospital-based cohort study. Lancet Infect Dis. 2021;21:1246–56.

51. Hirotsu Y, Omata M. Discovery of a SARS-CoV-2 variant from the P.1 lineage harboring K417T/E484K/N501Y mutations in Kofu, Japan. J Infect. 2021;82:276–316.

52. Planas D, Bruel T, Grzelak L, Guivel-Benhassine F, Staropoli I, Porrot F, et al. Sensitivity of infectious SARS-CoV-2 B.1.1.7 and B.1.351 variants to neutralizing antibodies. Nat Med. 2021;27:917–24.

53. Poland GA, Ovsyannikova IG, Kennedy RB. SARS-CoV-2 immunity: review and applications to phase 3 vaccine candidates. Lancet (London, England). 2020;396:1595–606.

54. Pollet J, Chen W-H, Strych U. Recombinant protein vaccines, a proven approach against coronavirus pandemics. Adv Drug Deliv Rev. 2021;170:71–82.

55. Oliveira SC, de Magalhães MTQ, Homan EJ. Immunoinformatic analysis of SARS-CoV-2 nucleocapsid protein and identification of COVID-19 vaccine targets. Front Immunol. 2020;11:587615.

56. Bonam SR, Kotla NG, Bohara RA, Rochev Y, Webster TJ, Bayry J. Potential immuno-nanomedicine strategies to fight COVID-19 like pulmonary infections. Nano Today. 2021;36:101051.

57. Radzikowska U, Ding M, Tan G, Zhakparov D, Peng Y, Wawrzyniak P, et al. Distribution of ACE2, CD147, CD26, and other SARS-CoV-2 associated molecules in tissues and immune cells in health and in asthma, COPD, obesity, hypertension, and COVID-19 risk factors. Allergy Eur J Allergy Clin Immunol. 2020;75:2829–45.

58. Hikmet F, Méar L, Edvinsson Å, Micke P, Uhlén M, Lindskog C. The protein expression profile of ACE2 in human tissues. Mol Syst Biol. 2020;16:e9610.

59. Singh M, Bansal V, Feschotte C. A single-cell RNA expression map of human coronavirus entry factors. Cell Rep. 2020;32:108175.

60. Wicik Z, Eyileten C, Jakubik D, Simões SN, Martins DC, Pavão R, et al. ACE2 interaction networks in COVID-19: a physiological framework for prediction of outcome in patients with cardiovascular risk factors. J Clin Med. 2020;9:3743.

61. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis. 1997;18:2714–23.

62. Krieger E, Joo K, Lee J, Lee J, Raman S, Thompson J, et al. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: four approaches that performed well in CASP8. Proteins Struct Funct Bioinforma. 2009;77:114–22.

63. Ordog R. PyDeT, a PyMOL plug-in for visualizing geometric concepts around proteins. Bioinformation. 2008;2:346–7.

64. Tina KG, Bhadra R, Srinivasan NPIC. Protein Interactions Calculator. Nucleic Acids Res. 2007;35:W473–6.

65. Lobley A, Sadowski MI, Jones DT. pGenTHREADER and pDomTHREADER: new methods for improved protein fold recognition and superfamily discrimination. Bioinformatics. 2009;25:1761–7.

66. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER suite: protein structure and function prediction. Nat Methods. 2014;12:7–8.

67. Trisolini L, Gambacorta N, Gorgoglione R, Montaruli M, Laera L, Colella F, et al. FAD/NADH dependent oxidoreductases: from different amino acid sequences to similar protein shapes for playing an ancient function. J Clin Med. 2019;8:2117 (**MDPI AG**).

68. Tragni V, Cotugno P, De Grassi A, Cavalluzzi MM, Mincuzzi A, Lentini G, et al. Targeting Penicillium expansum GMC oxidoreductase with high affinity small molecules for reducing patulin production. Biology (Basel). 2020;10:21.

69. Pierri CL, Bossis F, Punzi G, De Grassi A, Cetrone M, Parisi G, et al. Molecular modeling of antibodies for the treatment of

TNFα-related immunological diseases. Pharmacol Res Perspect. 2016;4:e00197.

70. Bossis F, De Grassi A, Palese LLL, Pierri CLL. Prediction of high- and low-affinity quinol-analogue-binding sites in the aa3 and bo3 terminal oxidases from Bacillus subtilis and Escherichia coli1. Biochem J. 2014;461:305–14.

71. Pierri CL, Parisi G, Porcelli V. Computational approaches for protein function prediction: A combined strategy from multiple sequence alignment to molecular docking-based virtual screening. Biochim Biophys Acta - Proteins Proteomics. 2010;1804:1695–712.

72. Persson B. Bioinformatics in protein analysis. Proteomics Funct Genomics. Basel: Birkhäuser Basel; 2000. p. 215–31.

73. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. Nucleic Acids Res. 2005;33:W382–8.

74. Van Durme J, Delgado J, Stricher F, Serrano L, Schymkowitz J, Rousseau F. A graphical interface for the FoldX forcefield. Bioinformatics. 2011;27:1711–2.

75. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;580–5. https://doi.org/10.1038/ng.2653.

76. Yuan M, Wu NC, Zhu X, Lee C-CD, So RTYY, Lv H, et al. A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. Science. 2020;368:630–3.

77 Palmieri F, Pierri CL. Structure and function of mitochondrial carriers - role of the transmembrane helix P and G residues in the gating and transport mechanism. FEBS Lett. 2010;584:1931–9.

78. Ballesteros JA, Deupi X, Olivella M, Haaksma EE, Pardo L. Serine and threonine residues bend alpha-helices in the chi(1) = g(-) conformation. Biophys J. 2000;79:2754–60.

79. Li F, Li W, Farzan M, Harrison SC. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Science. 2005;309:1864–8.

80. Cozier GE, Arendse LB, Schwager SL, Sturrock ED, Acharya KR. Molecular basis for multiple omapatrilat binding sites within the ACE C-domain: implications for drug design. J Med Chem. 2018;61:10141–51.

81. Ray K, Hines CS, Coll-Rodriguez J, Rodgers DW. Crystal structure of human thimet oligopeptidase provides insight into substrate recognition, regulation, and localization. J Biol Chem. 2004;279:20480–9.

82. Brown CK, Madauss K, Lian W, Beck MR, Tolbert WD, Rodgers DW. Structure of neurolysin reveals a deep channel that limits substrate access. Proc Natl Acad Sci U S A. 2001;98:3127–32.

83. Reva BA, Finkelstein AV, Skolnick J. What is the probability of a chance prediction of a protein structure with an rmsd of 6 Å? Fold Des. 1998;3:141–7.

84. Walls AC, Tortorici MA, Snijder J, Xiong X, Bosch BJ, Rey FA, et al. Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. Proc Natl Acad Sci U S A. 2017;114:11157–62.

85. Alejandra Tortorici M, Walls AC, Lang Y, Wang C, Li Z, Koerhuis D, et al. Structural basis for human coronavirus attachment to sialic acid receptors. Nat Struct Mol Biol. 2019;26:481–9.

86. Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell. 2020;181:281-292.e6.

87. Pinto D, Park YJ, Beltramello M, Walls AC, Tortorici MA, Bianchi S, et al. Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. Nature. 2020;583:290–5.

88. Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature. 2020;581:215–20.

89. Deupi X, Olivella M, Govaerts C, Ballesteros JA, Campillo M, Pardo L. Ser and Thr residues modulate the conformation of pro-kinked transmembrane alpha-helices. Biophys J. 2004;86:105–15.

90. Tian F, Tong B, Sun L, Shi S, Zheng B, Wang Z, et al. N501y mutation of spike protein in sars-cov-2 strengthens its binding to receptor ace2. Elife. 2021;10:1–17.

91. Deshpande A, Harris BD, Martinez-Sobrido L, Kobie JJ, Walter MR. Epitope classification and RBD binding properties of neutralizing antibodies against SARS-CoV-2 variants of concern. Front Immunol. 2021;12:2185–99.

92. Lopez E, Haycroft ER, Adair A, Mordant FL, O'Neill M, Pymm P, et al. Simultaneous evaluation of antibodies that inhibit SARS-CoV-2 RBD variants with a novel competitive multiplex assay. JCI Insight. 2021;6:e150012.

93. Muik A, Wallisch A-K, Sänger B, Swanson KA, Mühl J, Chen W, et al. Neutralization of SARS-CoV-2 lineage B.1.1.7 pseudovirus by BNT162b2 vaccine-elicited human sera. Science. 2021;371:1152–3.

94. Doshi P. Covid-19: Do many people have pre-existing immunity? BMJ. 2020;370:m3563.

95. Ray JG, Schull MJ, Vermeulen MJ, Park AL. Association between ABO and Rh blood groups and SARS-CoV-2 infection or severe COVID-19 illness : a population-based cohort study. Ann Intern Med. 2021;174(3):308–15. https://doi.org/10.7326/m20-4511.

96. Tanaka S, Nelson G, Olson CA, Buzko O, Higashide W, Shin A, et al. An ACE2 Triple Decoy that neutralizes SARS-CoV-2 shows enhanced affinity for virus variants. Sci Rep. 2021;11:12740.

97. Warner FJ, Lew RA, Smith AI, Lambert DW, Hooper NM, Turner AJ. Angiotensin-converting enzyme 2 (ACE2), but not ACE, is preferentially localized to the apical surface of polarized kidney cells. J Biol Chem. 2005;280:39353–62.

98. Teixeira PF, Masuyer G, Pinho CM, Branca RMM, Kmiec B, Wallin C, et al. Mechanism of peptide binding and cleavage by the human mitochondrial peptidase neurolysin. J Mol Biol. 2018;430(3):348–62. https://doi.org/10.1016/j.jmb.2017.11.011.

99. Bestle D, Heindl MR, Limburg H, van Lam van T, Pilgram O, Moulton H, et al. TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. Life Sci Alliance. 2020;3:e202000786.

100. Filbin MR, Mehta A, Schneider AM, Kays KR, Guess JR, Gentili M, et al. Longitudinal proteomic analysis of severe COVID-19 reveals survival-associated signatures, tissue-specific cell death, and cell-cell interactions. Cell Reports Med. 2021;2:100287.

101 Kim SI, Pabon A, Swanson TA, Glucksman MJ. Regulation of cell-surface major histocompatibility complex class I expression by the endopeptidase EC3.4.24.15 (thimet oligopeptidase). Biochem J. 2003;375:111–20.

102. Portaro FCV, Gomes MD, Cabrera A, Fernandes BL, Silva CL, Ferro ES, et al. Thimet oligopeptidase and the stability of MHC class I epitopes in macrophage cytosol. Biochem Biophys Res Commun. 1999;255:596–601.

103. Dos Santos NB, Franco RD, Camarini R, Munhoz CD, Eichler RAS, Gewehr MCF, et al. Thimet oligopeptidase (EC 3.4.24.15) key functions suggested by knockout mice phenotype characterization. Biomolecules. 2019;9:382.

104. Boechat JL, Chora I, Morais A, Delgado L. The immune response to SARS-CoV-2 and COVID-19 immunopathology – current perspectives. Pulmonology. 2021;27:423–37.

105. Visniauskas B, Simões PSR, Dalio FM, Naffah-Mazzacoratti MDG, Oliveira V, Tufik S, et al. Sleep deprivation changes thimet oligopeptidase (THOP1) expression and activity in rat brain. Heliyon. 2019;5:e02896.

106. Stefano GB, Ptacek R, Ptackova H, Martin A, Kream RM. Selective neuronal mitochondrial targeting in SARS-CoV-2 infection affects cognitive processes to induce "Brain Fog" and results

in behavioral changes that favor viral survival. Med Sci Monit. 2021;e930886. https://doi.org/10.12659/msm.930886.

107. Shah VK, Firmal P, Alam A, Ganguly D, Chattopadhyay S. Overview of immune response during SARS-CoV-2 infection: lessons from the past. Front Immunol. 2020;11:1949.

108. Zhou T, Tsybovsky Y, Gorman J, Rapp M, Cerutti G, Chuang G-Y, et al. Cryo-EM Structures of SARS-CoV-2 Spike without and with ACE2 reveal a pH-dependent switch to mediate endosomal positioning of receptor-binding domains. Cell Host Microbe. 2020;28:867–79.

109. Omotuyi IO, Nash O, Ajiboye OB, Iwegbulam CG, Oyinloye BE, Oyedeji OA, et al. Atomistic simulation reveals structural mechanisms underlying D614G spike glycoprotein-enhanced fitness in SARS-COV-2. J Comput Chem. 2020;41:2158–61.

110. Mahmoudi Gomari M, Rostami N, Omidi-Ardali H, Arab SS. Insight into molecular characteristics of SARS-CoV-2 spike protein following D614G point mutation, a molecular dynamics study. J Biomol Struct Dyn. 2021;1–9. https://doi.org/10.1080/07391102.2021.1872418.

111. Skladany L, Koller T, Adamcova Selcanova S, Vnencakova J, Jancekova D, Durajova V, et al. Challenging management of severe chronic disorders in acute pandemic situation: Chronic liver disease under COVID-19 pandemic as the proof-of-principle model to orchestrate the measures in 3PM context. EPMA J. 2021;12:1–14.

112. Demerle K, Angus DCC, Seymour CWW. Precision Medicine for COVID-19: Phenotype Anarchy or Promise Realized? JAMA - J Am Med Assoc. 2021;2041–2. https://doi.org/10.1001/jama.2021.5248.

113. Zhou A, Sabatello M, Eyal G, Lee SSJ, Rowe JW, Stiles DF, et al. Is precision medicine relevant in the age of COVID-19?

Genet. Med. 2021. ;23(6):999–1000. https://doi.org/10.1038/s41436-020-01088-4.

114. Tonkin-Hill G, Ruybal-Pesántez S, Tiedje KE, Rougeron V, Duffy MF, Zakeri S, et al. Evolutionary analyses of the major variant surface antigen-encoding genes reveal population structure of Plasmodium falciparum within and between continents. PLoS Genet. 2021;17:e1009269.

115. De Grassi A, Caggese C, D'Elia D, Lanave C, Pesole G, Saccone C. Evolution of nuclearly encoded mitochondrial genes in Metazoa. Gene. 2005;354:181–8. https://doi.org/10.1016/j.gene.2005.03.046.

116. Adebali O, Reznik AO, Ory DS, Zhulin IB. Establishing the precise evolutionary history of a gene improves prediction of disease-causing missense mutations. Genet Med. 2016;18:1029–36.

117. Sanjuán R, Domingo-Calap P. Genetic diversity and evolution of viral populations. Encycl Virol. 2021;53–61. https://doi.org/10.1016%2FB978-0-12-809633-8.20958-8.

118. Irwin KK, Laurent S, Matuszewski S, Vuilleumier S, Ormond L, Shim H, et al. On the importance of skewed offspring distributions and background selection in virus population genetics. Heredity (Edinb). 2016;117:393–9.

119. Zhao X, Wang K, Zuo P, Liu Y, Zhang M, Xie S, et al. Early decrease in blood platelet count is associated with poor prognosis in COVID-19 patients—indications for predictive, preventive, and personalized medical approach. EPMA J. 2020;11(2):139–45.

Authors and Affiliations

Vincenzo Tragni[1] · Francesca Preziusi[1] · Luna Laera[1] · Angelo Onofrio[1] · Ivan Mercurio[1] · Simona Todisco[2] · Mariateresa Volpicella[1] · Anna De Grassi[1,3] · Ciro Leonardo Pierri[1,3]

✉ Ciro Leonardo Pierri
ciro.pierri@uniba.it; ciroleopierri1@gmail.com
https://browser-bioinf.com/

Anna De Grassi
https://browser-bioinf.com/

[1] Department of Biosciences, Biotechnologies, Biopharmaceutics, University of Bari, Via E. Orabona, 4, 70125 Bari, Italy

[2] Department of Sciences, University of Basilicata, Viale dell'Ateneo Lucano, 10-85100 Potenza, Italy

[3] BROWSer S.r.l. at Department of Biosciences, Biotechnologies, Biopharmaceutics, University "Aldo Moro" of Bari, Via E. Orabona, 4, 70126 Bari, Italy