Computational optimization of angiotensin-converting enzyme 2 for SARS-CoV-2 Spike molecular recognition

Lorenzo Di Rienzo [a,*], Michele Monti [b], Edoardo Milanetti [a,c], Mattia Miotto [a,c], Alberto Boffi [d], Gian Gaetano Tartaglia [a,b,e], Giancarlo Ruocco [a,c]

[a] Center for Life Nanoscience, Istituto Italiano di Tecnologia, Viale Regina Elena 291, 00161 Rome, Italy
[b] RNA System Biology Lab, Department of Neuroscience and Brain Technologies, Istituto Italiano di Tecnologia, Via Morego 30, 16163 Genoa, Italy
[c] Department of Physics, Sapienza University, Piazzale Aldo Moro 5, 00185 Rome, Italy
[d] Department of Biochemical Sciences "A. Rossi Fanelli", Sapienza University, Piazzale Aldo Moro 5, 00185 Rome, Italy
[e] Department of Biology and Biotechnology "Charles Darwin", Sapienza University, Piazzale Aldo Moro 5, 00185 Rome, Italy

A R T I C L E   I N F O

A B S T R A C T

Since the beginning of the Covid19 pandemic, many efforts have been devoted to identifying approaches to neutralize SARS-CoV-2 replication within the host cell. A promising strategy to block the infection consists of using a mutant of the human receptor angiotensin-converting enzyme 2 (ACE2) as a decoy to compete with endogenous ACE2 for the binding to the SARS-CoV-2 Spike protein, which decreases the ability of the virus to enter the host cell. Here, using a computational framework based on the 2D Zernike formalism we investigate details of the molecular binding and evaluate the changes in ACE2-Spike binding compatibility upon mutations occurring in the ACE2 side of the molecular interface. We demonstrate the efficacy of our method by comparing our results with experimental binding affinities changes upon ACE2 mutations, separating ones that increase or decrease binding affinity with an Area Under the ROC curve ranging from 0.66 to 0.93, depending on the magnitude of the effects analyzed. Importantly, the iteration of our approach leads to the identification of a set of ACE2 mutants characterized by an increased shape complementarity with Spike. We investigated the physico-chemical properties of these ACE2 mutants and propose them as bona fide candidates for Spike recognition.

© 2021 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creative-commons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection, caused by a new coronavirus emerged as a very infectious human pathogen, has rapidly spread all over the world and caused serious consequences for public health [1,2]. This coronavirus is related to bat coronaviruses and crossed into humans in late 2019 [3] where it causes large and heterogeneous symptoms such as fever and pneumonia, collectively called COVID-19 [2,4,5].

The entry of the virus in the human cells is mainly mediated by the Spike protein, a trimeric assembly that protrudes from the SARS-CoV-2 envelope [6–8]. Indeed, as repeatedly observed experimentally, each chain of the Spike trimer has a Receptor Binding Domain(RBD) able to bind the human protease angiotensin-converting enzyme 2 (ACE2), thus causing the Spike transition to

a new conformation that allows the membrane fusion and release of viral genetic material in the host cell cytosol [7,9–11].

The identification of possible molecular drugs able to prevent the Spike-ACE2 recognition is still an open problem. For instance, it has been demonstrated that monoclonal antibodies, targeting Spike protein, can represent an effective tool for the treatment of SARS-CoV-2 infection[12]. Moreover, the protective effect of Lactoferrin against SARS-CoV2 has been studied, even if the molecular mechanism of this action is still to be defined [13,14]. Another interesting approach is the one proposed by this recent work [15], where peptides mimicking ACE2 binding sites were designed in order to prevent the virus cell entry.

Among other strategies, soluble ACE2 (without transmembrane region) has been proposed as a possible drug, competing for the binding with the viral Spike protein and thus repressing the SARS-CoV-2 infection [16,17], since all the atomic contacts between Spike protein and ACE2 receptor occur in the extracellular domain of the human protease. Moreover, differently from the case of anti-Spike monoclonal antibodies recently identified and

---

* Corresponding author.
E-mail address: lorenzo.dirienzo@iit.it (L. Di Rienzo).

currently in clinical or pre-clinical stage [18–21], the eventual viral escape from this binding is very likely to consequently reduce also the affinity with the ACE2 receptor [22]. In this scenario, the identification of ACE2 mutants that increase the binding affinity with SARS-CoV-2 Spike protein can be extremely useful, both from diagnostic and therapeutic points of view [23].

Remarkably, in the recent work by Chan et al. [22], 117 sites on ACE2 sequence were identified, structurally sparse on the whole protein and each was mutated into all the other 19 possible natural residues. The resulting 2340 single mutations were analyzed in terms of affinity change with respect to the binding between wild type ACE2 and the Spike protein. Interestingly, several ACE2 mutations with improved binding to Spike RBD domain were identified, and an ACE2 engineered version with a combination of three of these single mutations was proposed.

However, if multiple simultaneous ACE2 residues mutations are considered, the amount of possible binding configurations dramatically increases, making the experimental exploration very difficult. An additional level of complexity is caused by the fact that Spike, as other parts of the virus, is subject to mutations that can affect the binding propensity to the corresponding receptor. Indeed, viral genomes undergo fast mutational processes that can produce variants with higher complementarity to the cell receptor and facilitate its inter-species transmission. As a matter of fact, even considering just one viral protein, the number of possible variants eventually occurring in its receptor-binding domain is very large. For these reasons, fast computational methods could be extremely useful to explore the space of possible mutations in terms of binding affinity to its receptor. In the last years, the computational molecular design has proven its importance and efficacy as a guide for optimizing molecular binding [24–27], which indicates that the strategy has a high chance for success.

Here we present a computational approach, based on the local shape characterization of protein surface through the 2D Zernike polynomial mathematical formalism, for the identification and the analysis of ACE2 mutants characterized by an increased geometrical complementarity with SARS-CoV-2 Spike protein. In recent years the 3D Zernike approach has been applied to study the compatibility of interacting molecular surfaces [28–31] proving to be a powerful, albeit computationally expensive, tool. We employed the 2D Zernike formalism to evaluate the shape complementarity of protein–protein interfaces and in particular the Spike-ACE2 one [32]. Indeed, Using the 2D formalism allows us to gain an order of magnitude advantage in terms of computational time and no significant loss in description accuracy, since the complementarity calculated with 3D or 2D formalism highly correlate (see Supplementary Fig. 1).

Once a patch of the molecular surface is extracted, its points can be properly projected on a plane and the region shape can be represented as a 2D function and then expanded on the basis of the Zernike polynomials. This allows us to summarize the geometrical properties of a protein region in an ordered set of numerical descriptors invariant under rotation.

In this work, we first performed a computational mutagenesis approach reproducing all the interfaces mutations evaluated in the work of Chan et al. [22], demonstrating that the shape complementarity measure obtained with our Zernike formalism achieves a very good agreement with experimental results. Thanks to the low computational cost of the protocol for protein–protein compatibility evaluation upon mutation, independently of the number of substituted residues, we next modeled and ranked also all the possible double amino-acid mutations, evaluating a number of mutants largely inaccessible for any experimental technique. Finally, combining our approach with a coarse-grained technique for the evaluation of electrostatics at the interface, we devised a general and widely applicable algorithm for the identification of

the set of ACE2 mutants characterized by a very high compatibility with Spike RBD and specific physico-chemical features.

Remarkably, these designed mutants can represent interesting candidates for a soluble inhibitor for SARS-CoV-2 infection, competing with membrane-anchored ACE2 for Spike binding.

2. Methods and theory

2.1. Surface construction

The experimental structure of RBD-Spike complex we used is the one labeled with the PDB code 6vw1 [33]: it was solved with X-ray crystallography with a resolution of 2.68 A. The structure report ACE2 residues from Serine 19 to Alanine 614 and RBD residues from Asparagine 334 to Proline 527. Starting from the experimental structure, the computational mutagenesis has been performed using the SCWRL4 software [34].

Atomic charges and radii were assigned using PDB2PQR [35]. Solvent Accessible Surface and are computed using dms software [36].

2.2. Interface representation and complementarity evaluation

We defined the ACE2 and Spike binding sites as the set of residues closer than 5 Å to any atom of the molecular partner. As evident in Fig. 1A, the interaction between these two proteins Spike and ACE2 receptor is mediated by two separate molecular regions (region A and region B).

Given the structure of the experimental molecular complex (PDB: 6vw1), we extracted the surface of the interacting regions (Spike A, Spike B, ACE2 A, ACE2 B) from the whole surface and, using the procedure we developed in [32], we represent these patch surfaces as 2D functions (see Fig. 1B). We then expanded the patch surfaces in the 2D Zernike polynomial basis, compactly summarizing their geometrical shape in an ordered and invariant set of numerical descriptors. Since the shape of two perfectly fitting surfaces is the same in the Zernicke formalism, adopting a pairwise metric between the descriptors we can efficiently measure the complementarity between corresponding regions in the ACE2-Spike interface (Spike A – ACE2 A, Spike B – ACE2 B, where the lower is the distance the higher is the complementarity). By defining S_A, S_B, A_A, A_B as the vector of Zernike descriptors regarding Spike region A, Spike region B, ACE2 region A and ACE2 region B respectively, the shape complementarity is defined as:

$$C = \frac{d(S_A, A_A) + d(S_B, A_B)}{2} \tag{1}$$

where d(X, Y) represents the euclidean distance between two vectors.

To evaluate the long-range Coulomb contributions to the binding compatibility, we adopted a Coarse Grained atomic representation. After assigning to each atom its partial charge at physiological pH, we schematize each residue with a main chain bead and a side chain bead. These beads, given their corresponding atoms, are located at their geometrical center and are charged with the sum of their partial charges. The coordinates and the charges of the beads regarding the interface residues of both Spike and ACE2 are available in Supporting Information (Supplementary Tables 1 and 2). In this way we can compute the total interface Coulomb Energy, summing up all the contributions of the beads closer than 10 Å to beads of the molecular partner (See Fig. 1C). In particular, selected the $i = 1, \ldots, N$ ACE2 interface beads and the $j = 1, \ldots, M$ Spike interface beads, we define the electrostatic interface energy:

$$E_{int} = \sum_{i=1}^{N} \sum_{j=i}^{M} E_{ij} = \sum_{i=1}^{N} \sum_{j=1}^{M} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} \tag{2}$$
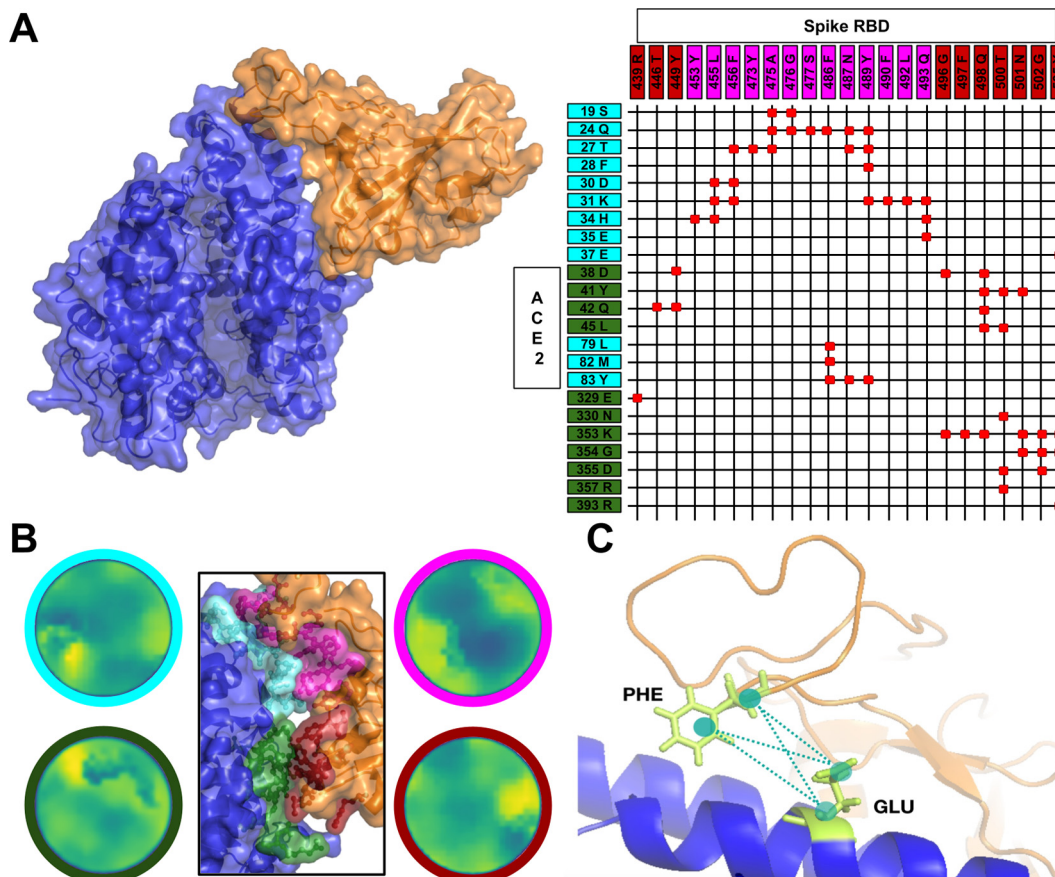
Fig. 1. Surface complementarity and electrostatic evaluation A) Molecular surface and cartoon representation of the ACE2-Spike RBD complex (PDB id:6vw1). Residues found in structural proximity (nearer than 5 Å) are highlighted by red dots in the matrix on the right. The molecular contact between these 2 proteins occurs through 2 different regions, therefore we defined 4 set of residues: Spike A (453 Y, 455 L, 456 F, 473 Y, 475 A, 476 G, 477 S, 486 F, 487 N, 489 Y, 490 F, 492 L, 493 Q) colored in purple, ACE2 A (19 S, 24 Q, 27 T, 28 F, 30 D, 31 K, 34 H, 35 E, 37 E, 79 L, 82 M, 83 Y) colored in cyan, Spike B (439 R, 446 T, 449 Y, 496G, 497 F, 498 Q, 500 T, 501 N, 502 G, 505 Y) colored in red, ACE2 B (38 D, 41 Y, 42 Q, 45 L, 329 E, 330 N, 353 K, 354 G, 355 D, 357 R, 393 R) colored in green. As shown in the matrix, residues of ACE2 A (cyan) interact only with ones of Spike A (purple), as well as ACE2 B (green) contacts only Spike B (red). B) Zernike disks associated with the interaction region ACE2 A(enclosed in cyan), ACE2 B (green), Spike A (purple), Spike B (red). In the disks the palette ranges from yellow (low distance from observation point) to green (high distance from the observation point). In the center the atomic details of the interactions are reported, where the interacting regions are shown with the corresponding color. C) Coarse-Grained representation of a couple of interacting surface residues. Each residue is associated with two beads, one in place of main chain atoms and another for side chain ones (see Methods).

where q_i and q_j are the bead charges, r_{ij} is the distance between the beads and ϵ_0 is the permittivity. Note that the favorable energy is achieved when E_{int} is negative.

2.3. 2D Zernike polynomials and invariants

Given a function in a 2D space defined in the unit circle, $f(r, \phi)$, it can be expanded in the Zernike polynomials basis as:

$$f(r, \phi) = \sum_{n=0}^{\infty} \sum_{m=0}^{m=n} c_{nm} Z_{nm} \tag{3}$$

where

$$c_{nm} = \frac{(n+1)}{\pi} \langle Z_{nm} | f \rangle = \frac{(n+1)}{\pi} \int_0^1 dr r \int_0^{2\pi} d\phi Z_{nm}^*(r, \phi) f(r, \phi). \tag{4}$$

being the expansion coefficients, while the complex functions, $Z_{nm}(r, \phi)$ are the Zernike polynomials.

Since the modulus of each coefficient ($z_{nm} = |c_{nm}|$) does not depend on the phase, the zernike description is invariant for rotations around the origin of the unitary circle. Therefore the shape complementarity between any two patches, disregarding their size or orientation, can be evaluated using the Zernike descriptors of their 2D projections.

In particular, we assessed the complementarity between patch i and j as the euclidean distance between the invariant vectors:

$$d_{ij} = \sqrt{\sum_{k=1}^{M=121} (z_i^k - z_j^k)^2} \tag{5}$$

All the procedure for the calculation and the comparison of the Zernike invariants is made using an *in-house* python code, available on GitHub at: https://github.com/matmi8/Zernike2D.

2.4. Chemical physical properties of the mutants

A number of properties, including hydrophobicity, α-helical, β-sheet, disorder, burial, aggregation, membrane, and nucleic acid-binding propensities, are employed to build physico-chemical 'profiles' of ACE2 mutants. For each mutant, the profile is calculated directly from the primary structure, where we have chosen 80 scales that, for each amino acid, furnishes a numerical value quantifying a given feature. [37]. The Physico-chemical properties are compared to identify similarities and differences.

To compute how a mutated sequence have changed with respect to the wt one, we defined the formula (8), and we normalized the obtained values using Z-score.

Following the CleverMachine approach [37,38], we also check consistencies among predictors of the same Physico-chemical propensity by grouping them and generating a 'consensus' (Fig. 5B). The used properties can be divided into 8 macro categories corresponding to the following properties: Aggregation, Disorder, Membrane Propensity, Hydrophobicity, Burial, Nucleic Acid Binding, α Helix, β sheet. Since each scale takes a different approach to compute its score, to give an overall characterization of the resulting sequence, we average the X_{diff} over all the scale belonging to the same macro group.

3. Results and discussions

In this work, we extensively studied the effects of mutations of ACE2 residues on the ACE2-Spike interface. While carrying out residue substitutions, we assessed the interface shape complementarity and the electrostatic compatibility of the mutated molecular complex.

We first tested our modeling approach on the available experimental data by evaluating shape and electrostatics changes caused by previously studied mutations. Comparing our results with the ones presented in the recent paper by Chan et al. [22], our formalism recognize with an Area under the ROC curve of 0.66 the favorable mutations from unfavorable ones (taking in consideration the mutations with the greatest effects ROC AUC increase up to 0.93). Because of the high accuracy of our method, we set up an iterative mutation process that leads to the computational design of ACE2 soluble variants optimized for strong binding to SARS-CoV-2 Spike protein. We analyzed the physico-chemical properties of the mutants and observed that the increased molecular complementarity is accompanied by a decreased propensity to form α-helix, promoting hydrophobicity and tendency to aggregate.

The computational modeling we adopted for the analysis of the ACE2-Spike binding interface is summarized in Fig. 1. Once identified the interacting fraction of molecular surfaces, we calculated their Zernike descriptors obtaining a set of numerical values describing the shape of proteins binding regions. Simply calculating the Euclidean distance between two sets of descriptors we estimate the complementarity between the corresponding molecular regions. Moreover, we developed a coarse-grained representation for the evaluation of the electrostatic compatibility between these two proteins (See Methods section for details).

3.1. Mutations and changes in binding affinity

We studied the effect of several ACE2 residues substitutions on Spike interfaces.

For each mutation, we performed the mutagenesis, obtaining a new ACE2 structure (see Methods section). We assessed the compatibility of the mutated version with SARS-CoV-2 Spike, in terms of both shape and electrostatics. The shape complementarity is studied by evaluating the distance between the Zernike descriptors of the mutated structure and the ones of the Spike protein, determining if the mutation has caused an increase or a decrease of the shape complementarity (Eq. (1)). In parallel, we evaluated the electrostatic compatibility analyzing the electrostatic interface energy produced by mutated arrangements (Eq. (2)). The overall shape complementarity balance is:

$$\Delta C = C_{wt} - C_{mut} \tag{6}$$

where C is the shape complementarity defined in Eq. (1) and the subscripts wt and mut refer to the complementarity obtained with the wild type and mutated ACE2. When two interfaces are characterized by an high complementarity the C value calculated between their Zernike descriptors is low: therefore when $\Delta C > 0$ the

mutation is favourable. Similarly the electrostatics energy balance can be written as:

$$\Delta E = E_{wt} - E_{mut} \tag{7}$$

where E represent the interface electrostatic energy calculated using Eq. (2), obtained in the wild type or mutated interface. A favorable mutation results in a $\Delta E > 0$.

To test the reliability of our procedure we compared the results of our computational approach with the experimental ones obtained by [22], focusing only on the mutation involving the ACE2 binding site residues. Indeed, since the Zernike approach deals with shape complementarity at the interface, we selected the set of ACE2 residues closer than 5 Å to any atom of Spike residues, assuming that their mutation can significantly alter the ACE2 binding site shape. In other words, we focused on 23 ACE2 sequence positions that correspond to a total number of 437 point amino acid substitutions since each residue can be mutated in each of the 19 other residues.

In Fig. 2 we reported the results of this study, focusing on shape complementarity. For each mutation we have both the experimental and the computational score (calculated using Eq. (6)): ordering the mutations according to their experimental scores and selecting the top and bottom N%, we built a binary classifier based on the computational scores. Performing a Receiver Operating Characteristics (ROC) curve analysis, we show in Fig. 2A the Area Under the ROC curve (AUC) as a function of the considered top (highest experimental scores) and bottom (lowest experimental scores) cases (N%).

Examining the most diverse cases, i.e. the mutations characterized by a very high increase and decrease in binding affinity, we found that the Zernike computational protocol distinguishes very well favorable from deleterious mutations. Indeed, for a low fraction of the dataset (from 8% to 2% of the dataset considered) the AUC ranges in high values (the best value we obtained is 0.93 considering 2% top and bottom of the dataset). Increasing N the AUC decreases, notwithstanding stabilizing its value to 0.66 (N = 50%, whole dataset), a satisfactory result considering that the computational prediction of residues mutation effects, in terms of binding affinity of protein–protein interaction, is still an open problem (especially when no machine-learning approaches are used)[39]. To further highlight the reliability of the procedure in identifying the favorable/deleterious mutations, in panels B and C we report the ROC curves and the distributions of shape Zernike scores when are considered the top and bottom 8% and 50% mutations, respectively. A detailed description of the statistical comparison between our results and experimental data can be found in Supporting Information. In Supplementary Fig. 2 we show the correlation between experimental and computational score, highlighting as, at least in terms of mean behavior, the variations in terms of shape complementarity correlate with the experimental variations in binding affinity upon mutations. Moreover, in Supplementary Fig. 3, we show the distribution of the ΔE obtained for these point mutations. As evident, the most probable ΔE value is around 0 kcal/mol (60% of mutations have an effect between −2 and 2 kcal/mol), meaning that most of the mutations do not cause a remarkable effect on the electrostatic pairing between ACE2 and Spike. As one could expect, only a minority of mutations (9%) causes a significant energetic gaining ($\Delta E > 2 Kcal/mol$) in respect to the wild type, while a larger fraction of mutants worse the electrostatic compatibility ($\Delta E < -2 Kcal/mol$). In Supplementary Fig. 4 we report the AUC of the ROC curve as well the AUC of the Precision-Recall curve and the global accuracy we get as a function of the fraction of the dataset each time considered. In Supplementary Table 3 we report the 10 top and 10 bottom mutations, ranked according to their experimental score of variation in Spike-ACE2
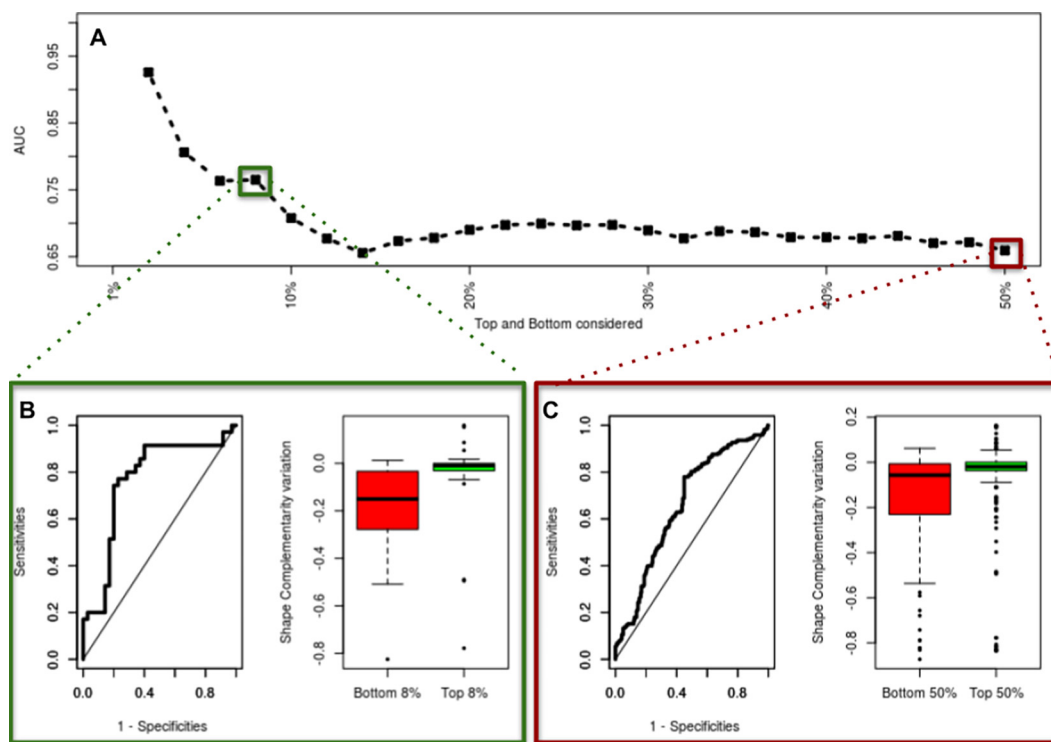
Fig. 2. Agreement between experimental and computational scores of binding affinity changes upon mutation. A) Area under the Receiving Operating Characteristics (ROC) curve of the classifier employing the computational Zernike scores of variations in shape complementarity, as a function of the fraction of cases considered. Note that N% means (i.e. 10%) that we defined 2 groups of cases selecting the N% of mutations from top and bottom experimental changes in binding affinity (e.g. taking N = 50% means dealing with 100% of the dataset, simply dividing it in two groups). The experimental data are taken from [22]. B) and C) ROC curves and computational scores boxplot distributions regarding the cases with N = 8% and N = 50% respectively.

binding affinity, in order to underline that such mutations are not localized in some specific positions.

These analyses confirm that our computational modeling is effective in capturing the main aspects of the effect of the mutations on the binding.

3.2. Design of multiple ACE2 mutants increasing SARS-CoV-2 Spike complementarity

The described computational framework evaluates the effects of an ACE2 residue substitution, in terms of binding compatibility with SARS-CoV-2 Spike, in a very limited amount of time and with good agreement with experimental data. In principle, it is therefore possible to push the complementarity assessment to the level of multiple mutations, evaluating the consequences of a number of mutations inaccessible to any experimental technique, due to their cost in terms of time and money.

We modeled all the possible double mutations obtainable from the 23 ACE2 interface residues, and we analyzed the results obtained in terms of Zernike descriptors. We note that the mutants proposed with this procedure are over 91000, which indicates the large-scale possibility of our protocol. We initially investigated to what extent the effects of double mutations can be described as the sum of the 2 single mutations composing it. Indeed, in Fig. 3 we plotted the sum of the effects of 2 mutations when considered independently as a function of the effect of the corresponding double mutation. Not surprisingly, the vast majority of the points lies on the diagonal line meaning that the consequences of the mutations are substantially independent. Fitting the observed value to the diagonal straight line, we calculated the residual for each point, that is the difference between the observed value and the expected one. In our case, since the estimated trend is simply y = x, the

residual for each point is the difference between the Zernike scoring obtained considering jointly two mutations and the sum of the two mutations independent scoring. In Supplementary Fig. 5 we report the values and cumulative distributions of such residuals. In absolute values, 95% of the points have a residual lower than 0.102 or, in other words, 95% of the scatter points falls between the 2 dashed lines flanking the diagonal. The points in Fig. 3 are colored according to the distance between the C_α of the involved residue: as expected, the points that deviate from the general trend are the ones corresponding to a couple of very close residues, since their combined mutations modify the conformations observed individual case. To quantify this effect, defining as "positive" the cases with double mutation score higher than 0 and as "negative" the cases with Double mutation score lower than 0, we can consider the sum of the Zernike score of the single mutations as the "predictor" (See Fig. 3). With this framework, we obtained a sensitivity of 95.6%, a specificity of 98.9%, and precision of 92.9%.

It is interesting to discuss the upper left corner points, forming a second linear trend. These points are only 367 (out of over 91000) characterized by residual higher than 0.7, representing therefore the points more distant from the diagonal. Analyzing the couples of residues involved in these double mutations, interestingly it turns out that 355 (97% to all these points) are due to the simultaneous mutation of Asp 30 and Lys 31. These are oppositely charged, large and consecutive residues: therefore it is expectable that the effect of their combined substitution has to be considered cooperatively. In particular, when only one of these 2 residues is singularly mutated is very unlikely to obtain an improvement in shape complementarity: when, on the contrary, these 2 residues are jointly mutated can be caused a overall positive effect on the compatibility with Spike.

In light of these considerations, we developed an algorithm for the design of ACE2 mutants characterized by compatible interface
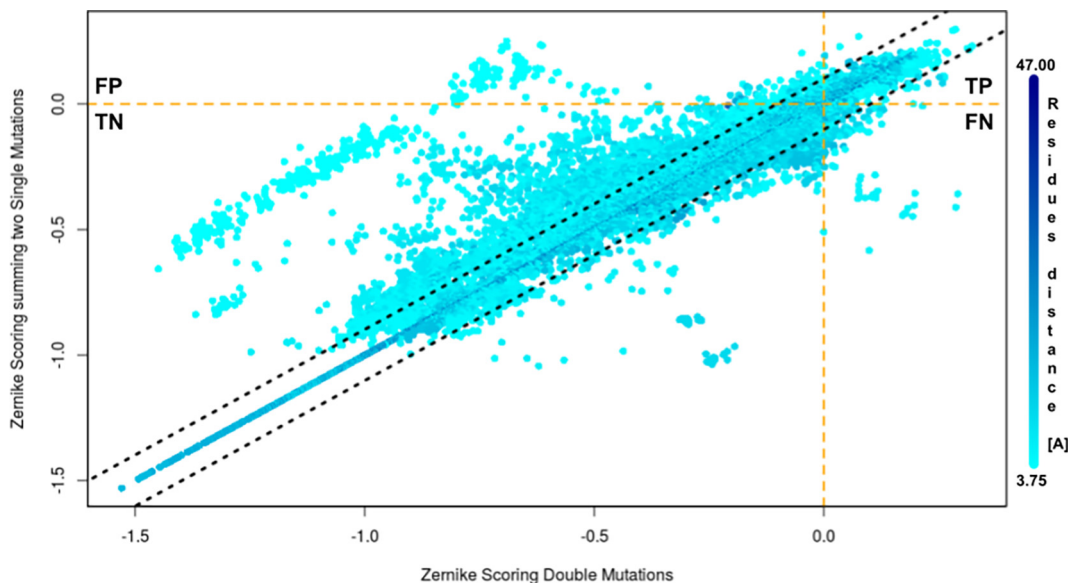
Fig. 3. Comparison between the Zernike scores obtained after a double mutation or combining two single mutation scores. Dots are colored from cyan to dark blue as the distance between the considered residues increases. The black lines enclose the 95% of the points, those characterized by deviation from to the straight line y = x lower than 0.1. The TP quadrant includes couples of mutations that increase the shape complementarity if they are considered both independently or combined. The FP quadrant included couples of mutations that increase shape complementarity if they are considered separately, while the combined effect on the contrary worsen it. The TN quadrant includes couples of mutations that decrease the shape complementarity if they are considered both independently or combined. The FN quadrant included couples of mutations that decrease shape complementarity if they are considered separately, while the combined effect on the contrary increases it.

electrostatics and by a notably increased shape complementarity (see Fig. 4A). In particular, starting from the wild type ACE2 sequence, we modeled all the possible point mutations. Filtering only the mutations characterized by non-deleterious electrostatics ($\Delta E > 0$), we selected the best two mutations in terms of shape complementarity (max(ΔC)). Established these mutations as a starting point for a new iteration, we repeated twice the mutation protocol on the remaining binding site residues, obtaining 4 double site substitutions. Performing on 5 levels of point mutations, we identified 32 possible ACE2 mutants characterized by increased shape complementarity and compatible electrostatics. For a single case, in which we selected each time the best mutation, we continued the mutation protocol until all the binding site residues are mutated: interestingly, after 5 mutations the shape complementarity reaches a plateau value, meaning that the inclusion of other substitutions does not advantage the ACE2-Spike interaction (Sup-

plementary Fig. 6). For this reason, we decided to deal with 5 residues mutants.

The shape complementarity gaining of the mutants built with this protocol are shown in Fig. 4B. Note that, since the distance between wild type Zernike descriptors is 1.48 and the best gaining in shape complementarity is 0.45, we improve the shape complementarity by over 30% with respect to the wild type. Analyzing the sum of the effect of the single mutations as a function of the combined effects of the 5 position mutations (Supplementary Fig. 7), we note that the majority of the points lies in the proximity of the diagonal, testifying to substantial independence of the various mutation consequences. In Fig. 4C, we reported the most frequent mutations we obtained: it is worth noting that, to gain complementarity with Spike, the residues H34, L79, K31, N330 and L45 result to be mutated in more than 60% of the proposed mutants.
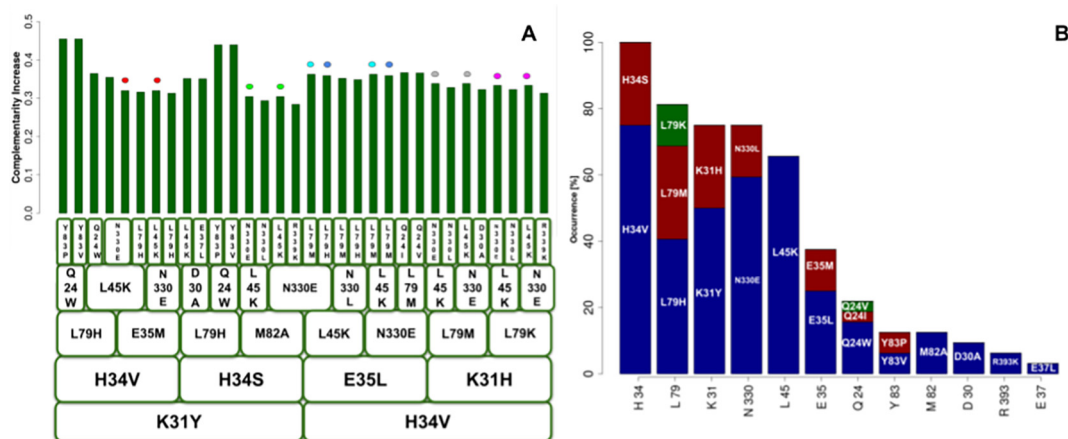

Fig. 4. Outcomes of the mutational protocol for ACE2 optimization. A) Schematic representation of the mutational protocol: starting from the wild type form of ACE2, all possible single mutations are explored and the two best variants in terms of shape complementarity and electrostatic energy are selected. The process is then iterated, each time starting from the selected variants, ending with 32 novel versions of ACE2. The bars represent the shape complementarity gaining of the mutants built with the protocol. B) Frequencies of the mutations observed in the mutagenesis protocol.

3.3. Physico-chemical characterization of mutants

In this section, we analyze the properties of the ACE2 sequences we found with the optimization protocol we described above, in particular computing the physico-chemical characteristics changes of the mutant with respect to the wild type.

We used 80 amino acid property scales, describing features ranging from amino acid hydrophobicity to the probability to be found in a given secondary structure or involved in aggregation, as defined in a previous work [37]. In these scales, each of the 20 amino acids is characterized by a numerical value describing its propensity to a specific feature. These scales can be grouped into 8 categories, according to the described amino acid tendency (the propensity of aggregation, disorder, to be in the membrane, hydrophobicity, to be a buried residue, nucleic acid binding, α-helix, and β-sheet). We preliminary correlate the scales belonging to the same categories (See Supplementary Fig. 8): since we aim to obtain a stable characterization, we removed from the analysis the scales marked by a mean correlation with the other scales of the same category lower than 0.562 (p-value 0.01). Considering the remaining scales, therefore, each position of ACE2 sequences,

wild type or mutants, was substituted with the value of the corresponding residue: in order to consider even the surrounding effects, each sequence position value in the final description is obtained averaging on the 7 residue interval centered on it.

In this framework, aiming to a compact quantification of the extent of changes occurring between mutated and wild type sequence, we defined the following formula:

$$X_{diff} = \frac{n_0}{N} \sum_{1=1}^{N} \frac{X_i^{mut} - X_i^{wt}}{X_i^{wt}} \qquad (8)$$

where i indicates the position on the ACE2 sequence of length N, n_0 is the number of sequence positions experiencing changes in physico-chemical properties upon mutation, and X is a generic scale. In other words, X_{diff} represent the variation in percentages caused by the amino acid substitutions concerning the examined property: therefore we have a X_{diff} for each of 32 mutants and for each of the remaining 56 amino acid tendency.

Grouping these scales according to the described characteristics [37] and performing a Z-score normalization, we obtained the colormap reported in Fig. 5A (the colormap before the normalization
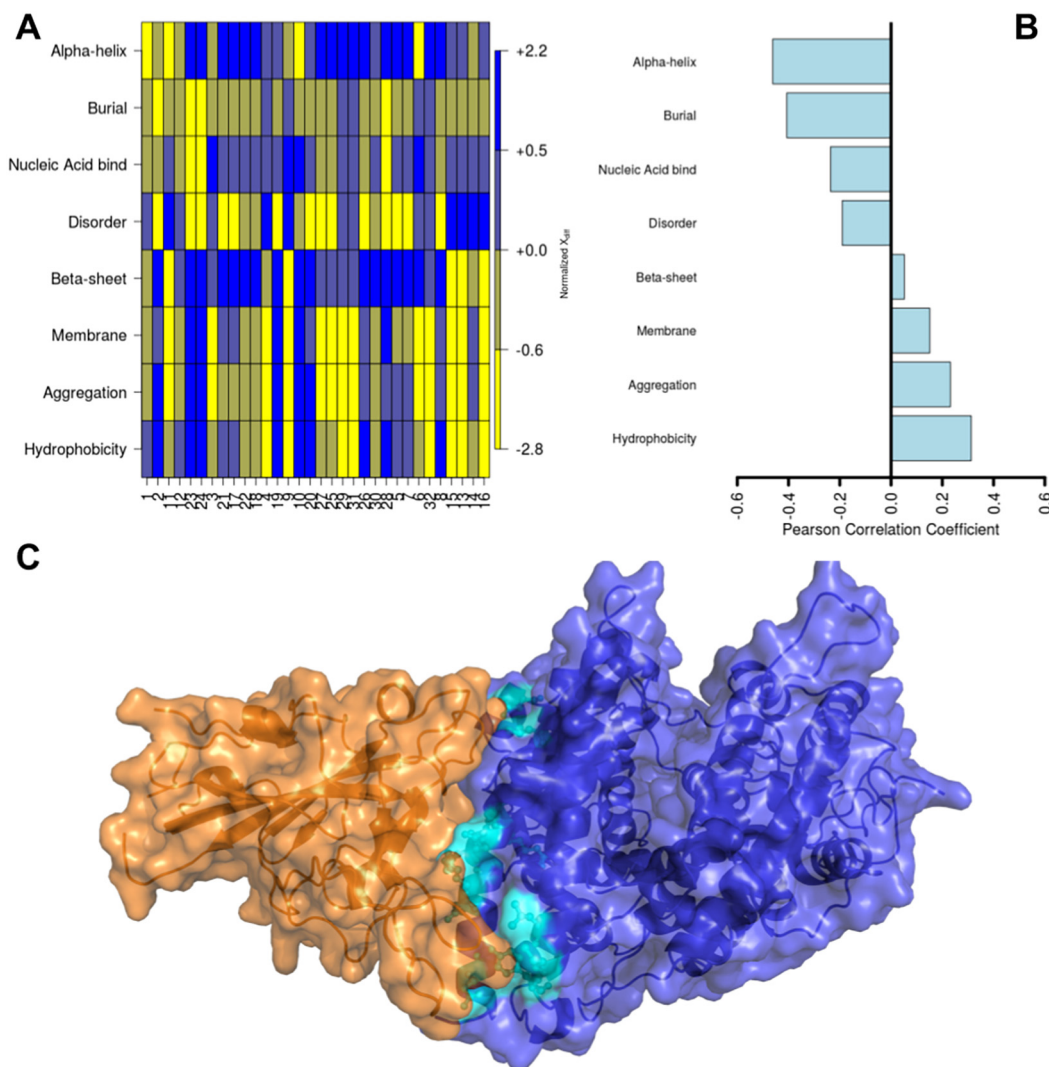

Fig. 5. Analysis of the physico-chemical properties of the 32 proposed ACE2 variants. A) Colormap of the X_{diff} descriptors for each of the 32 mutants we identified with our procedure and for each of the 8 macro-characteristics analyzed. B) Pearson Correlation values of the 8 mean descriptors X_{diff} with the mutants shape complementarity gaining in terms of Zernike descriptors. C) Molecular representation of the interface between ACE2 (blue) and Spike (orange). The positions that have undergone a mutation in our procedure are highlighted in cyan.

is provided in Supplementary Fig. 9): the color represents the extent of the variation, X_{diff}, for each mutant we identified and for each property. The mutants in this colormap are ordered according to the gaining in shape complementarity according to Zernike formalism.

The correlations between the Zernike scores and the change in the analyzed properties are shown in the barplot in Fig. 5B. Interestingly, the main result of this analysis is that the best mutations identified, for which shape complementarity to Spike is increased, are characterized by a decrease in α-helix secondary structure propensy and increase in hydrophobicity. This result acquires significance if we consider that the dominant secondary structure of ACE2 protein is α-helical(See Fig. 5C): the mutants we identified are characterized by the tendency to slightly alter the structure of ACE2 binding site in favour of tighter interactions. This result is in agreement with previous reports indicating that the hydrophobicity and aggregation propensities of interfaces are key characteristics that promote formation of stable contacts [40,41]. In Supplementary Table 4 we report the X_{diff} we obtained when all the single mutations we encounter are performed alone.

In order to evaluate the effects of such mutations on the overall protein secondary structure stability, we predict for both the wild type and the mutants sequences the secondary structure using NetSurfP WebServer [42]. Analogously, we perform a graph analysis, where each residue represent a node in order to underline that we work on peripheral residues and that their mutations should not interfere with the global protein fold. The results of our analysis (See Supplementary Figs. 10 and 11) testify that the residue substitution occurs on peripheral residues and the caused physico-chemical modifications are slightly enough to not constitute a danger for protein secondary and tertiary stability.

4. Conclusions

Current protein and mRNA vaccines require the expression of SARS-CoV-2 Spike protein in the human cell. After Spike is reproduced in the cell, its presence causes a T-lymphocytes and B-lymphocytes response that protects the cell from future SARS-CoV-2 infection. Several monoclonal antibodies have been recently validated and more are under investigation for antiviral clinical use [12,43–45].

An alternative approach is to use a 'mock' ACE2 without transmembrane region to compete for the binding to Spike when infection occurs [16,17], making them unavailable for binding with cell membrane-anchored ACE2 and cell invasion.

To effectively adopt such a strategy, in addition to further studies needed to understand how strengthening the solubility of the protein to guarantee a stronger concentration and, therefore, better response in the host cell, it is fundamental the mutants ACE2 affinity for Spike is increased [22].

Indeed, through an innovative approach we recently developed that exploits Zernike's formalism [32], we developed a computational protocol for the design of mutants that increase the binding propensity of ACE2 for Spike.

Given the very high number of mutations we modeled in this work, after each substitution we can not fully relax the structure since it would require a too time-consuming minimization procedure: for this reason, we have to evaluate the interface electrostatic compatibility with a coarse-grained approach, since the computational mutagenesis procedure could turn out in atom–atom interaction unphysically too close and energetic. Indeed, even if the size of the ACE2-RBD system is affordable using a full-atomic model, for example classical molecular dynamics have produced many interesting results [46], we would need to perform a molecular simulation for each of the tens of thousands of the mutations

considered, after whom we can proceed to the full-atomic electrostatic judgment. This would require an unaffordable computational cost. In addition, it is worth noting that in our work we deal with only static models, and therefore the importance of well-documented dynamics effects in ACE2-RBD recognition [47,48,46,49,50] can not be fully captured by our protocol.

This notwithstanding, the efficacy of such formalism in understanding the main effects of residues mutations in Spike-ACE2 compatibility is demonstrated comparing our results with a large experimental campaign recently conducted [22], achieving a ROC AUC ranging from 0.93 to 0.66 according to the magnitude of experimental binding affinity changes considered.

Performing iterative cycles of single mutations upon ACE2 binding site, our protocol identifies a set of mutants characterized with an increased shape complementarity and compatible electrostatics with Spike binding region. Indeed, even if when the protein interface undergoes a considerable number of mutations the possibility of errors due to methodology limits obviously increases, the main advantage of our protocol is the possibility to evaluate with a low computational cost the combined effects of the possible mutations, making judgeable a very large volume in the space of the possible interface configurations.

Interestingly, the selected residues substitutions are correlated with an increase in hydrophobicity, which indicates augmentation of the propensity to form more stable interactions, as proposed in previous works [40,41].

We envisage that these mutants could represent a promising starting point for the identification of SARS-CoV-2 Spike inhibitors.

CRediT authorship contribution statement

Lorenzo Di Rienzo: Investigation, Conceptualization, Methodology, Software. **Michele Monti:** Methodology, Software, Conceptualization. **Edoardo Milanetti:** Conceptualization, Methodology. **Mattia Miotto:** Conceptualization, Methodology. **Alberto Boffi:** Conceptualization. **Gian Gaetano Tartaglia:** Conceptualization, Methodology. **Giancarlo Ruocco:** Conceptualization, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.csbj.2021.05.016.

References

[1] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, et al. Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. The Lancet 2020;395(10223):497–506.

[2] Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, et al. A novel coronavirus from patients with pneumonia in china. New Engl J Med 2019.

[3] Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. The Lancet 2020;395(10223):470–3.

[4] Wang W, Tang J, Wei F. Updated understanding of the outbreak of 2019 novel coronavirus (2019-ncov) in wuhan, china. J Med Virol 2020;92(4):441–7.

[5] Ellul M, Benjamin L, Singh B, Lant S, Michael B, Kneen R, et al. Neurological associations of covid-19. Available at SSRN 3589350..

[6] Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, Si H-R, Zhu Y, Li B, Huang C-L, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020:1–4.

[7] 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..

[8] Wan Y, Shang J, Graham R, Baric RS, Li F. Receptor recognition by novel coronavirus from wuhan: An analysis based on decade-long structural studies of sars. J Virol..

[9] Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al., Sars-cov-2 cell entry depends on ace2 and tmprss2 and is blocked by a clinically proven protease inhibitor, Cell..

[10] Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of the sars-cov-2 by full-length human ace2. Science..

[11] Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, Graham BS, McLellan JS. Cryo-em structure of the 2019-ncov spike in the prefusion conformation. Science 2020;367(6483):1260–3.

[12] Barnes CO, Jette CA, Abernathy ME, Dam K-MA, Esswein SR, Gristick HB, Malyutin AG, Sharaf NG, Huey-Tubman KE, Lee YE, et al. Sars-cov-2 neutralizing antibody structures inform therapeutic strategies. Nature 2020:1–6.

[13] Salaris C, Scarpa M, Elli M, Bertolini A, Guglielmetti S, Pregliasco F, Blandizzi C, Brun P, Castagliuolo I. Protective effects of lactoferrin against sars-cov-2 infection in vitro. Nutrients 2021;13(2):328.

[14] Miotto M, Di Rienzo L, Bò L, Boffi A, Ruocco G, Milanetti E. Molecular mechanisms behind anti sars-cov-2 action of lactoferrin. Front Mol Biosci 8..

[15] Freitas FC, Ferreira PHB, Favaro DC, Oliveira RJ d. Shedding light on the inhibitory mechanisms of sars-cov-1/cov-2 spike proteins by ace2-designed peptides. J Chem Inf Model 2021;61(3):1226–43.

[16] Hofmann H, Geier M, Marzi A, Krumbiegel M, Peipp M, Fey GH, Gramberg T, Pöhlmann S. Susceptibility to sars coronavirus s protein-driven infection correlates with expression of angiotensin converting enzyme 2 and infection can be blocked by soluble receptor. Biochem Biophys Res Commun 2004;319 (4):1216–21.

[17] Monteil V, Kwon H, Prado P, Hagelkrüys A, Wimmer RA, Stahl M, et al., Inhibition of sars-cov-2 infections in engineered human tissues using clinical-grade soluble human ace2. Cell..

[18] Hansen J, Baum A, Pascal KE, Russo V, Giordano S, Wloga E, Fulton BO, Yan Y, Koon K, Patel K, et al. Studies in humanized mice and convalescent humans yield a sars-cov-2 antibody cocktail. Science 2020;369(6506):1010–4.

[19] Wec AZ, Wrapp D, Herbert AS, Maurer DP, Haslwanter D, Sakharkar M, Jangra RK, Dieterle ME, Lilov A, Huang D, et al. Broad neutralization of sars-related viruses by human monoclonal antibodies. Science 2020;369(6504):731–6.

[20] Wang C, Li W, Drabek D, Okba NM, van Haperen R, Osterhaus AD, van Kuppeveld FJ, Haagmans BL, Grosveld F, Bosch B-J. A human monoclonal antibody blocking sars-cov-2 infection. Nat Commun 2020;11(1):1–6.

[21] Rogers TF, Zhao F, Huang D, Beutler N, Burns A, He W-T, Limbo O, Smith C, Song G, Woehl J, et al. Isolation of potent sars-cov-2 neutralizing antibodies and protection from disease in a small animal model. Science 2020;369 (6506):956–63.

[22] Chan KK, Dorosky D, Sharma P, Abbasi SA, Dye JM, Kranz DM, Herbert AS, Procko E. Engineering human ace2 to optimize binding to the spike protein of sars coronavirus 2. Science 2020;369(6508):1261–5.

[23] Zoufaly A, Poglitsch M, Aberle JH, Hoepler W, Seitz T, Traugott M, Grieb A, Pawelka E, Laferl H, Wenisch C, et al. Human recombinant soluble ace2 in severe covid-19. Lancet Respir Med 2020;8(11):1154–8.

[24] Whitehead TA, Chevalier A, Song Y, Dreyfus C, Fleishman SJ, De Mattos C, Myers CA, Kamisetty H, Blair P, Wilson IA, et al. Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. Nat Biotechnol 2012;30(6):543–8.

[25] Choi Y, Furlon JM, Amos RB, Griswold KE, Bailey-Kellogg C. Disruppi: structure-based computational redesign algorithm for protein binding disruption. Bioinformatics 2018;34(13):i245–53.

[26] Liu Y, Kuhlman B. Rosettadesign server for protein design. Nucl Acids Res 2006;34(suppl_2):W235–8..

[27] Di Rienzo L, Milanetti E, Testi C, Montemiglio LC, Baiocco P, Boffi A, Ruocco G. A novel strategy for molecular interfaces optimization: the case of ferritin-transferrin receptor interaction. Comput Struct Biotechnol J 2020;18:2678–86.

[28] Di Rienzo L, Milanetti E, Alba J, D'Abramo M. Quantitative characterization of binding pockets and binding complementarity by means of zernike descriptors. J Chem Inf Model 2020;60(3):1390–8.

[29] Sandomenico A, Di Rienzo L, Calvanese L, Iaccarino E, D'Auria G, Falcigno L, Chambery A, Russo R, Franzoso G, Tornatore L, et al. Insights into the interaction mechanism of dtp3 with mkk7 by using std-nmr and computational approaches. Biomedicines 2021;9(1):20.

[30] Kihara D, Sael L, Chikhi R, Esquivel-Rodriguez J. Molecular surface representation using 3d zernike descriptors for protein shape comparison and docking. Curr Protein Peptide Sci 2011;12(6):520–30.

[31] Venkatraman V, Yang YD, Sael L, Kihara D. Protein-protein docking using region-based 3d zernike descriptors. BMC Bioinf 2009;10(1):407.

[32] Milanetti E, Miotto M, Di Rienzo L, Monti M, Gosti G, Ruocco G. 2d zernike polynomial expansion: finding the protein-protein binding regions. Comput Struct Biotechnol J 2020;19:29.

[33] Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, Geng Q, Auerbach A, Li F. Structural basis of receptor recognition by sars-cov-2. Nature 2020;581 (7807):221–4.

[34] Krivov GG, Shapovalov MV, Dunbrack Jr RL. Improved prediction of protein side-chain conformations with scwrl4. Proteins Struct Funct Bioinf 2009;77 (4):778–95.

[35] Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. Pdb2pqr: an automated pipeline for the setup of poisson–boltzmann electrostatics calculations. Nucl Acids Res 2004;32(suppl_2):W665–7..

[36] Richards FM. Areas, volumes, packing, and protein structure. Annu Rev Biophys Bioeng 1977;6(1):151–76.

[37] Klus P, Bolognesi B, Agostini F, Marchese D, Zanzoni A, Tartaglia GG. The cleversuite approach for protein characterization: predictions of structural properties, solubility, chaperone requirements and rna-binding abilities. Bioinformatics 2014;30(11):1601–8.

[38] Klus P, Ponti RD, Livi CM, Tartaglia GG. Protein aggregation, structural disorder and rna-binding ability: a new approach for physico-chemical and gene ontology classification of multiple datasets. BMC Genom 2015;16(1):1–7.

[39] Geng C, Xue LC, Roel-Touris J, Bonvin AM. Finding the $\delta\delta g$ spot: Are predictors of binding affinity changes upon mutations in protein–protein interactions ready for it?. Wiley Interdiscip Rev Comput Mol Sci 2019;9(5):e1410.

[40] Pechmann S, Levy ED, Tartaglia GG, Vendruscolo M. Physicochemical principles that regulate the competition between functional and dysfunctional association of proteins. Proc Natl Acad Sci 2009;106 (25):10159–64.

[41] Masino L, Nicastro G, Calder L, Vendruscolo M, Pastore A. Functional interactions as a survival strategy against abnormal aggregation. FASEB J 2011;25(1):45–54.

[42] Petersen B, Petersen TN, Andersen P, Nielsen M, Lundegaard C. A generic method for assignment of reliability scores applied to solvent accessibility predictions. BMC Struct Biol 2009;9(1):1–10.

[43] Paiva KJ, Grisson RD, Chan PA, Huard RC, Caliendo AM, Lonks JR, King E, Tang EW, Pytel-Parenteau DL, Nam GH, et al. Validation and performance comparison of three sars-cov-2 antibody assays. J Med Virol 2021;93 (2):916–23.

[44] Weisberg SP, Connors TJ, Zhu Y, Baldwin MR, Lin W-H, Wontakal S, Szabo PA, Wells SB, Dogra P, Gray J, et al. Distinct antibody responses to sars-cov-2 in children and adults across the covid-19 clinical spectrum. Nat Immunol 2021;22(1):25–31.

[45] Wang Z, Schmidt F, Weisblum Y, Muecksch F, Barnes CO, Finkin S, Schaefer-Babajew D, Cipolla M, Gaebler C, Lieberman JA, et al. mrna vaccine-elicited antibodies to sars-cov-2 and circulating variants. Nature 2021:1–7.

[46] Francés-Monerris A, Hognon C, Miclot T, García-Iriepa C, Iriepa I, Terenzi A, Grandemange S, Barone G, Marazzi M, Monari A. Molecular basis of sars-cov-2 infection and rational design of potential antiviral agents: modeling and simulation approaches. J Proteome Res 2020;19(11):4291–315.

[47] Bernardi A, Huang Y, Harris B, Xiong Y, Nandi S, McDonald KA, Faller R. Development and simulation of fully glycosylated molecular models of ace2-fc fusion proteins and their interaction with the sars-cov-2 spike protein binding domain. PloS One 2020;15(8):e0237295.

[48] Spinello A, Saltalamacchia A, Magistrato A. Is the rigidity of sars-cov-2 spike receptor-binding motif the hallmark for its enhanced infectivity? Insights from all-atom simulations. J Phys Chem Lett 2020;11(12):4785–90.

[49] Williams-Noonan BJ, Todorova N, Kulkarni K, Aguilar M-I, Yarovsky I. An active site inhibitor induces conformational penalties for ace2 recognition by the spike protein of sars-cov-2. J Phys Chem B 2021;125(10):2533–50.

[50] Liu D, Chen X, Long D. Nmr-derived conformational ensemble of state 1 of activated ras reveals insights into a druggable pocket. J Phys Chem Lett 2020;11(9):3642–6.