THE JOURNAL OF PHYSICAL CHEMISTRY Letters

Physical Insights into the Biosphere, Atmosphere, and Space

The Answer Lies in the Energy: How Simple Atomistic Molecular Dynamics Simulations may Hold the Key to Epitope Prediction on the Fully Glycosylated SARS-CoV-2 Spike Protein

Stefano Artin Serapian, Filippo Marchetti, Alice Triveri, Giulia Morra, Massimiliano Meli, Elisabetta Moroni, Giuseppe A. Sautto, Andrea Rasola, and Giorgio Colombo

J. Phys. Chem. Lett., **Just Accepted Manuscript** • DOI: 10.1021/acs.jpclett.0c02341 • Publication Date (Web): 04 Sep 2020

The answer lies in the energy: how simple atomistic molecular dynamics simulations may hold the key to epitope prediction on the fully glycosylated SARS-CoV-2 spike protein

Stefano Serapian[1,†] Filippo Marchetti[1,2,†] Alice Triveri[1], Giulia Morra[3], Massimiliano Meli[3], Elisabetta Moroni[3], Giuseppe A. Sautto[4], Andrea Rasola[5], Giorgio Colombo[1,*]

1) University of Pavia, Department of Chemistry, viale Taramelli 12, 27100 Pavia (Italy)
2) University of Milan, Department of Chemistry, via Venezian 21, 20133 Milano (Italy)
3) SCITEC-CNR, via Mario Bianco 9, 20131 Milano (Italy)
4) Center for Vaccines and Immunology, Department of Infectious Diseases, University of Georgia, 501 D.W. Brooks Drive, 30602 Athens - GA (USA)
5) Dipartimento di Scienze Biomediche, Università di Padova, viale G. Colombo 3, 35131 Padova (Italy)

†) These authors contributed equally

*) Corresponding author: g.colombo@unipv.it

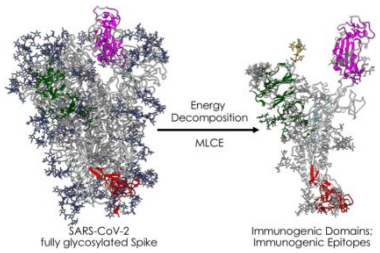
Abstract

SARS-CoV-2 is a health threat with dire socioeconomical consequences. As the crucial mediator of infection, the viral glycosylated Spike protein (S) has attracted the most attention and is at the center of efforts to develop therapeutics and diagnostics.

Herein, we use an original decomposition approach to identify energetically uncoupled substructures as antibody binding sites on the fully glycosylated S. Crucially, all that is required are unbiased MD simulations; no prior knowledge of binding properties or ad-hoc parameter combination is needed.

Our results are validated by experimentally confirmed structures of S in complex with anti- or nanobodies. We identify poorly coupled subdomains, poised to host (several) epitopes, and potentially involved in large functional conformational transitions. Moreover, we detect two distinct behaviors for glycans: those with stronger energetic coupling are structurally relevant and protect underlying peptidic epitopes; those with weaker coupling could themselves be prone to antibody recognition.

TOC Graphic



Keywords: SARS-CoV-2; Immune Reactivity; Epitope Discovery; Structural Vaccinology; Molecular Dynamics

The novel coronavirus SARS-CoV-2, the etiological agent of COVID-19 respiratory disease, has infected millions of people worldwide, causing more than 800000 deaths (as of August 30th, 2020) and extensive social and economic disruption. Given the pandemic status of the outbreak, social distancing measures cannot be sufficient any longer in containing it on a worldwide scale. This emergency calls for the development of strategies to rapidly identify pharmacological agents or vaccines as the only way to contain and combat the disease, restoring normal social conditions. Indeed, a number of currently ongoing trials focus on developing vaccines (see for instance https://www.nytimes.com/interactive/2020/science/coronavirus-vaccine-tracker.html) or on repurposing drugs already developed for other disorders [1-4].

SARS-CoV-2 is extraordinarily effective in exploiting the host's protein machinery for replication and spreading. This is a characteristic that it shares with other members of the *Coronaviridae* family, all of which are characterized by a highly selective tropism that determines the onset of a variety of diseases in domestic and wild animals as well as in humans, including central nervous system affections, hepatitis and respiratory syndromes [5-6]. As was the case with its human predecessors SARS-CoV and MERS, the homotrimeric viral Spike protein (S) (Figure 1) is the key player regulating cell entry (with the protein receptor angiotensin-converting enzyme 2 (ACE2) representing the host cell docking point in SARS-CoV-2 and SARS-CoV) [7-8]. The CoV S protein is then cleaved by a series of serine proteases, including trypsin, cathepsins, elastase, the host type 2 transmembrane serine protease (TMPRSS2), and plasmin, which promote virus entry into epithelial cells [4]. In this

context, it is important to underline that many vaccines under development for SARS-CoV-2 indeed focus on using recombinant forms of the S protein.

Recent cryo-electron microscopy (cryo-EM) analyses allowed to precisely determine the structure of the full-length Spike protein in its trimeric form [9-11], and the structural basis for the recognition of the Spike protein's Receptor Binding Domain (RBD; Figure 1) by the extracellular peptidase domain of ACE2 [7]. In parallel, computational studies have started to provide atomically detailed insights into S protein dynamics and on the elaborate role of the diverse polysaccharide chains that decorate its surface, effectively shielding a large portion of it from the host [12-14]. Computational approaches have also started to shed light on the determinants of binding to host cell receptors, studying in particular the interactions of the S protein with ACE2 [15-17].

This detailed dynamic and structural knowledge can set the stage for understanding the molecular bases of S protein recognition by the host's immune system, providing information on which physico-chemical determinants are required to elicit functional antibodies. Such understanding can then be exploited to design and engineer improved antigens based on S, for instance by identifying antigenic domains that can be expressed in isolation or short sequences (epitopes) that can be mimicked by synthetic peptides[18-23]: this would be a crucial first step in the selection and optimization of candidate vaccines and therapeutic antibodies (on top of those already in development), as well as in the development of additional serological diagnostic tools.

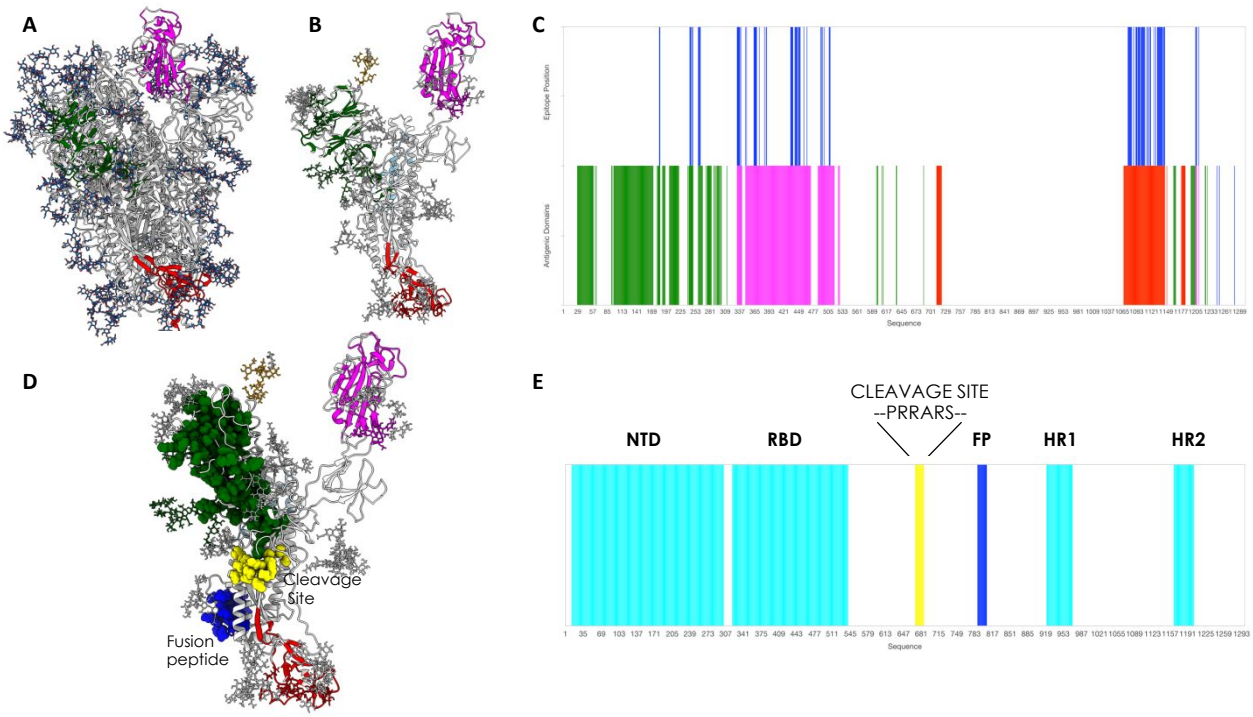


Figure 1. 3D structure, glycosylation and location of antigenic domains and epitopes on SARS-CoV-2 fully glycosylated Spike protein. A) The starting, fully glycosylated Spike protein trimer. The coating oligosaccharides are colored in dark blue. The predicted antigenic domains are colored on the structure of one protomer. **B)** Isolated protomer with the most antigenic domains, detected via MLCE with the 15% cutoff, highlighted in colors: the antigenic part in the N-Domain is dark green; the part in the RBD is magenta; the part in the C-terminal domain is dark red. Oligosaccharides that define or are part of antigenic domains are also colored. Oligosaccharides that have a structural role and show strong energetic coupling to the protein are in white. **C)** The predicted antigenic sequences projected on the sequence of the protein. The bottom line reports the sequences defined as antigenic domains, with the same color code as in **B**. The top bar reports the location of peptidic epitopes identified with the most restrictive definition. **D)** Physical interaction between the boundaries of the predicted antigenic domain in the N-terminal region and the cleavage site of S. This subfigure also shows the physical proximity of the predicted C-terminal uncoupled region with the fusion peptide. **E)** Domain organization of the spike protein projected on the sequence. Numbering and domain definitions obtained from UNIPROT (https://www.uniprot.org/uniprot/P0DTC2).

Even more significantly, knowledge acquired today about such recognition mechanisms could well mean that we are better prepared to tackle similar pandemics in the future by contrasting them more efficiently through the application of the same efficient and well-tested methods to new protein variants. More specifically, upon emergence of a new

pathogen, generally portable computational methods could be advantageously exploited to rapidly identify and synthesize recombinant antigen or peptide-based vaccines.

Here, we analyze representative 3D conformations of the full-length trimeric S protein in its fully glycosylated form (Figure 1), extracted from atomistic molecular dynamics (MD) simulations provided by the Woods group [13, 24], in order to predict immunogenic regions. To this end, a simple ab initio epitope prediction method that we previously developed for unmodified proteins [25-29] is optimized and extended to cover glycoproteins. The method is based on the idea that antibody-recognition sites (epitopes) may correspond to localized regions only exhibiting low-intensity energetic coupling with the rest of the structure. Otherwise put, putative interacting patches are hypothesized to be characterized by non-optimized intramolecular interactions with the remainder of the protein. Actual binding to an external partner such as an Ab is expected to occur if favorable intermolecular interactions determine a lower free energy for the bound than the unbound state [25, 27, 30]. Furthermore, minimal energetic coupling with the rest of the protein provides these subregions with greater conformational freedom to adapt to and be recognized by a binding partner [28-29], as well as improved tolerance to mutations at minimal energetic expense without affecting the protein's native organization and stability in a way that could be detrimental for the pathogen: all these properties are indeed hallmarks of Ab-binding epitopes.

We show that our approach is indeed able to identify regions—also comprising carbohydrates—that recent structural immunology studies have shown to be effectively targeted by antibodies. On the same basis, our method predicts several additional potential

immunogenic regions (currently still unexplored) that can then be used for generating optimized antigens, either in the form of recombinant isolated domains or as synthetic peptide epitopes. Finally, our results help shed light on the mechanistic bases of the large-conformational changes underpinning biologically relevant functions of the protein.

To the best of our knowledge, this approach is one of the first that permits to discover epitopes in the presence of glycosylation (an aspect that is often overlooked), starting only from the analysis of the physico-chemical properties of the isolated antigen in solution. Importantly, the approach does not require any prior knowledge of antibody binding sites of related antigenic homologs and does not need to be trained/tuned with data sets or ad hoc combinations of information on sequences, structures, SASA or geometric descriptors. The procedure is thus immediately and fully portable to other antigens.

To reveal the regions of the S protein that could be involved in antibody (Ab) binding, we employ a combination of the Energy Decomposition (ED) and MLCE (Matrix of Low Coupling Energies) methods, which we previously introduced and validated[25-27, 31-39] and discuss in full in the Methods section.

Starting from 6 combined 400 ns replicas of atomistic molecular dynamics simulations of the fully glycosylated S protein in solution [13, 40] (built from PDB ID: 6VSB [9]), we isolate a representative frame from each of the three most populated clusters. ED and MLCE analyses of protein energetics assess the interactions that each aminoacid and glycan residue in S protomers establishes with every other single residue in the same protomer. In particular, we compute the nonbonded part of the potential energy (van der Waals, electrostatic interactions, solvent effects) implicitly, via an MM/GBSA calculation (molecular

mechanics/generalized Born and surface area continuum solvation)[41], obtaining, for a protomer composed of N residues (including monosaccharide residues on glycans), a symmetric $N \times N$ interaction matrix M_{ij}. Eigenvalue decomposition of M_{ij} highlights the regions of strongest and weakest coupling. The map of pairwise energy couplings can then be filtered with topological information (namely, the residue-residue contact map) to identify localized networks of low-intensity coupling (*i.e.*, clusters of spatially close residue pairs whose energetic coupling to the rest of the structure is weak and not energetically optimized through evolution).

In our model, when these fragments are located on or near the surface, contiguous in space and weakly coupled to the protein's 'stability core', they represent potential interaction regions (i.e., epitopes).

Once interacting vicinal residue pairs (i, j) are identified by cross-comparison with the residue-residue contact map (vide supra and Methods Section), identification of poorly coupled regions representing potential epitopes proceeds as follows. Residue pairs are firstly ranked in order of increasing interaction intensity (from weakest to strongest). Two distinct sets of energetically decoupled regions are then mapped by applying two distinct cutoffs ('softness thresholds') to the residue pair list: either from the first 15% or from the first 5% of the ranked pairs (i.e., the 5% or 15% of the residue pairs with the weakest energetic coupling). As a caveat, it is worth noting here that different S protomer conformations may provide slightly different results, as the interaction matrix is naturally dependent on the structural organization of the protein. Here, we use the combination of predicted energetically uncoupled sequences as the proposed immunoreactive domains or substructures.

The less restrictive 15% cutoff subdivides the full-length, fully folded S protein into potentially immunoreactive domains (see Figure 1B,C and Methods)[26, 30, 32]. The goal is to uncover regions that may normally be hidden from recognition by Abs in the native protein structure, but that can be experimentally expressed as isolated domains. Highly reactive neutralizing epitopes may in fact be present only in specific but transient conformations that are not immediately evident in the static X-ray and EM models of the protein or are not accessible even to large scale MD simulations. Presenting these (cryptic) regions for Ab binding through their isolated parent domains may prove more advantageous in developing new immunogens[26, 32].

The more stringent epitope definition (5% cutoff) narrows the focus on those (smaller) intra-domain regions that could be directly involved in forming the interface with Abs, and that can then be used to guide the engineering of optimized antigens in the form of synthetic epitope peptidomimetics. In this context, to be defined as epitopes, the energetically uncoupled regions must be at least 6-residue long.

Upon using the larger cutoff value, a large cluster of energetically unoptimized residue pairs localize at the Receptor Binding Domain, correctly identifying it as the most antigenic unit in the S protein's 'RBD up' protomer (Figure 1B,C magenta colored domain). Interestingly, when the lowest energy-coupled residue pairs are mapped onto the 'up' RBD of all three 3D structures isolated from MD, there is a large overlap with regions recognized by Abs and nanobodies (revealed by recent X-ray and cryo-EM structures). Importantly, for example, our calculation correctly identifies the binding region of mAb CR3022[42] (see pdb code

6w41.pdb), known to target a cryptic epitope that is exposed only upon significant structural rearrangement of the protein [12] (Figures 2 and Figure 4).

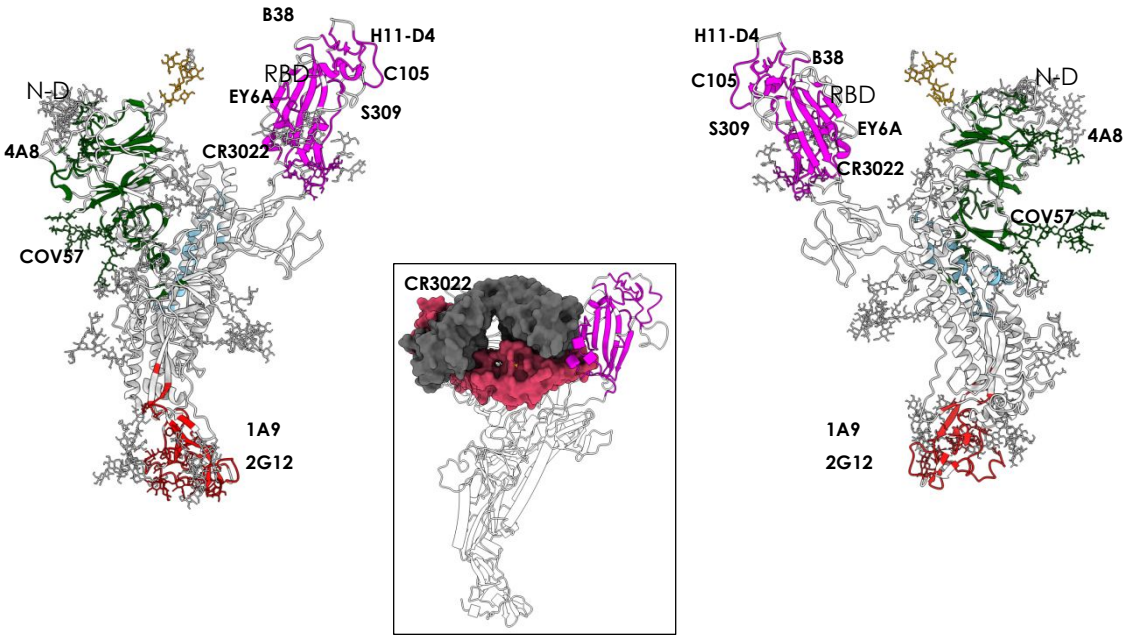


Figure 2. Antigenic domains and location of binding antibodies. The clusters of residues defining antigenic domains (dark green in the N-domain, magenta in the RBD, red in the C-terminal region) and the positions of the various antibodies whose structures and interactions in complexes with the full length protein have been described. The inset indicates the identification of the cryptic immunoreactive region binding CR3022.

A second domain that is found to host a large network of non-optimized interactions corresponds to the N-terminal domain (Figure 1B, C, D). The latter has been shown to bind the new antibody **4A8** (7c2l.pdb) in a paper that was published during the preparation of this manuscript [43].

A third region predicted to be highly antigenic coincides with the central/C-terminal part of the S1A domain. In a recent cryo-EM study of polyclonal antibodies binding to the S protein, this substructure was shown to be in the vicinity of the density for **COV57** Fab(s), a novel

Ab whose neutralizing activity showed no correlation with that of RBD-targeting Abs [44] (Figure 1B, D). We note here that MERS Ab 7D10 also binds in this region [45].

Furthermore, MLCE identifies a potentially highly reactive region in the S2 domain of the protein, in the CD region. This domain contains the epitope recently found to engage with 1A9 [46], an antibody recently shown to cross-react with S proteins of human, civet, and bat coronaviruses. This analysis also recognizes a potential antigenic region in a carbohydrate cluster located in the S2 domain of the protein: intriguingly, recent findings indicate that an oligosaccharide-containing epitope centered around this predicted region is targeted by the glycan-dependent antibody HIV-1 bnAb 2G12 [47] (Figure 1, 2).

Identification of energetically uncoupled domains also has mechanistic implications. Regions that are not involved in major intramolecular stabilization can be displaced from the biomolecule at minimal energetic costs, sustaining large-scale conformational changes that typically underpin its biological function. The boundary of the (uncoupled) N-terminal region (Figure 1, dark green domain) lies in physical proximity to the furin-targeted motif RRAR, which is essential for pre-activation of SARS-CoV-2 Spike through proteolysis. Thus, the large uncoupled region of the N-domain can synergize with (and favor, through domain displacement) cleavage of this motif, ultimately favoring detachment of S1-domain and release of the S2 fusion machinery[9-11, 48]. Furthermore, the beta-sheet at the initial boundary of the C-terminal domain in S2 (Red domain in Figure 1) is in close physical proximity to the fusion peptide (Figure 1D, E). Here, it would be reasonable to expect that exposure or conformational rearrangement of the C-terminal domain are favored by its non-optimized

interactions with the core of the S protein stalk, and would in turn optimally expose the fusion peptide favoring its integration with the host membrane [48].

Overall, these findings support the validity of our approach in identifying protein domains that can be aptly used as highly reactive immunogens, as they are most likely to be targeted by a humoral immune response. Our analysis predicts that regions other than the S protein RBD may represent alternative targets for neutralization or functional perturbation of SARS-CoV-2. On the one hand, this may be important considering the fact that RBD can also be the target on non-neutralizing antibodies, e.g. 3022 [42]. Indeed, using cocktails of antibodies to target different regions of S has recently been proposed as a viable therapeutic option [43], and Ab cocktails have been successfully used in the treatment of other epidemics such as Ebola. In this context, the company Regeneron is pursuing a cocktail type approach for SARS-CoV-2 that is already in clinical trials [49].

Turning to our more stringent definition of epitope, based exclusively on the top 5% of the most weakly coupled residue pairs (5% cutoff), we next focus on those regions of the S antigen that can be involved in forming contacts with antibodies.

Importantly, one predicted conformational epitope with sequence (348)A-(352)A-(375)S-(434)IAWNS(438)-(442)DSKVGG(447)-(449)YNYL(452)-(459)S-(465)E-(491)PLQS(494)-(496)Q-(507)PYR(509) encompasses regions of the S protein in contact with antibodies **C105** (6xcn.pdb)[44], **S309**(6wpt.pdb; 6wps.pdb) [50], **AB23** (7byr.pdb) [51]; with nanobody **H11-D4** (6z43.pdb); and with a recently reported synthetic nanobody (7c8v.pbd) (Figure 3).

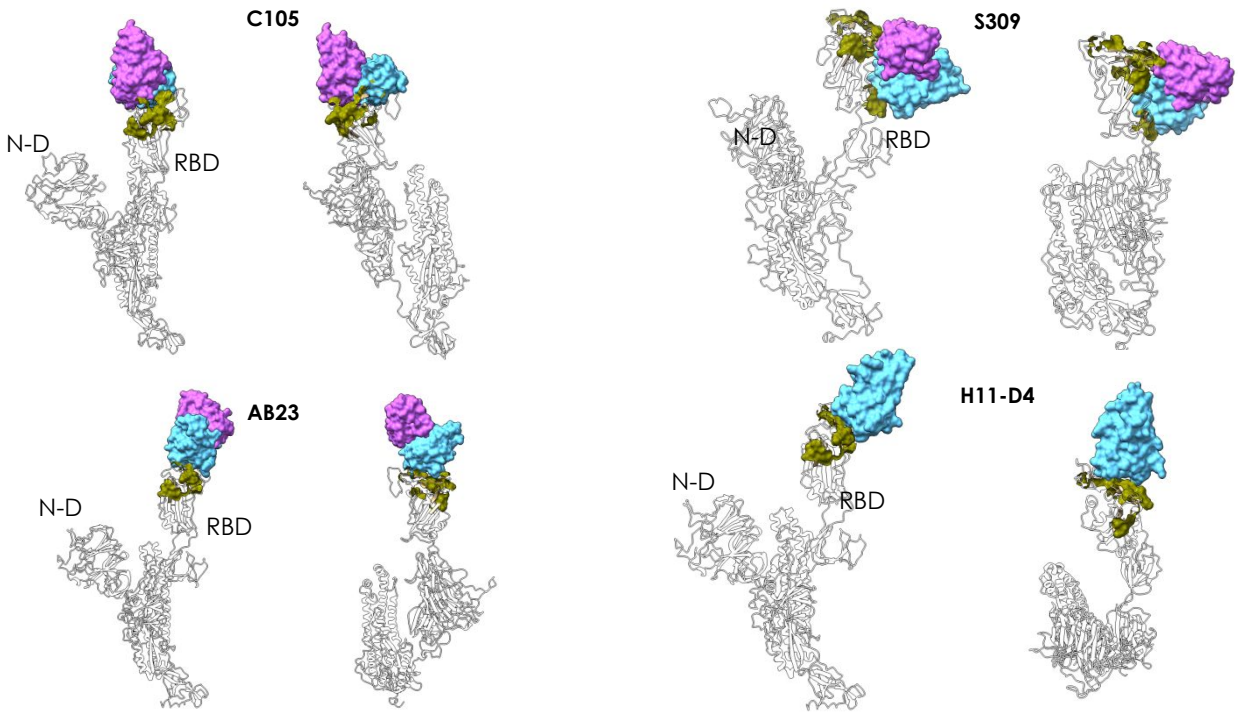


Figure 3. Peptidic epitopes predicted on the surface of the RBD using the restrictive definition of antigenic region and comparison with known Ab-complexes. The X-ray structures of the complexes between the various antibodies reported in the figure (C105, S309, AB23, and nanobody H11-D4) and the full length Spike protein are superimposed to the structure of the protomer used here for prediction. The green surfaces indicate the location of MLCE epitope predictions. The Fabs of the antibodies or of the nanobody are depicted as accessible surfaces in shades of blue.

Interestingly, an additional predicted patch comprising a set of decorating carbohydrates is correctly predicted to be part of the interface with antibody **S309 (6wpt.pdb; 6wps.pdb)** [50], with aminoacidic sequence (332)ITNLC(336)-(361)C and with the (*N*334-linked) fucosylated N-glycan chitobiose core (Manβ1-4GlcNAcβ1-4[Fucα1-6]GlcNAcβ-Asn) [52]. This predicted region sits notably close to the RBD interaction surface with ACE2.

Antibody **EY6A** (**6zdh.pdb**) binds the RBD in the region of the cryptic epitope described by Wilson and collaborators[42] (Figure 4).

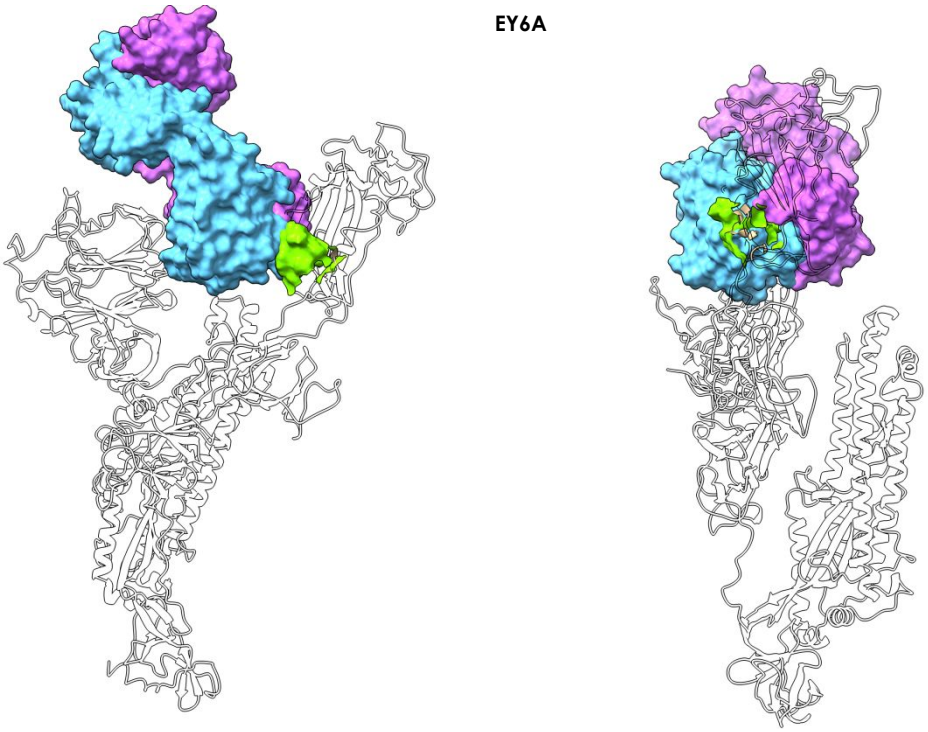


Figure 4. Antibody EY6A – Spike complex. The figure shows how Antibody **EY6A** (7byr.pdb) binds the RBD in the region of a cryptic epitope. The MLCE-predicted epitope region is shown in light green (lime) in two different orientations, indicating substantial contact formation with the antibody.

Importantly, our predicted patch (365)YSVLYN(370)-(384)PTKLN(388) covers a significant part of the epitope. Once again, it is worth remarking that identification of this potentially immunoreactive patch is simply and exclusively obtained from structural and energetic interaction data generated for a protomer of the glycosylated, isolated S protein, after unbiased MD simulation (see Methods section).

With the more restrictive epitope prediction cutoff we clearly identify a reactive area in the N-terminal domain of the Spike protein. The predicted patch (184)GN(185)-(242)LAL(244)-(246)R-(248)Y-(258)WTAGA(262) contains residues R246 and W258 which were described as central determinants for contact between the N-terminal domain and antibody **4A8** (7c2l.pdb) [43] (Figure 5).

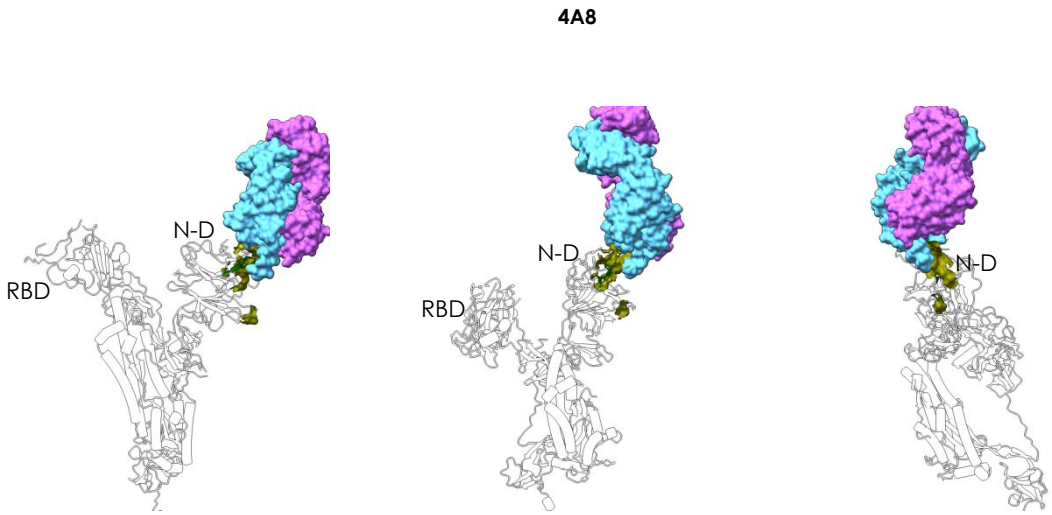


Figure 5. Antibody 4A8 – Spike complex. The figure shows how Antibody **4A8** binds the N-domain of Spike, supporting correct prediction of the epitope. The MLCE-predicted epitope region is shown in green in three different orientations, indicating substantial contact formation with the antibody. . The Fab of the antibody is depicted as accessible surface in shades of blue.

Finally, the restrictive prediction identifies the sequence spanning residues 1076-1146, which includes amino acids 1111-1130, experimentally identified as the epitope for the monoclonal mAb **1A9** [46]. Specifically our identified reactive sequence is the following: (1076) TTAPAICH(1083)-(1087)A-(1092)REG(1094)-(1096)FVSNGHWFVTQRN(1108)-(1112)P-(1114)I-(1116)T-(1118)DN(1119)-(1126)C-(1129)V-(1132)IVNNTVYDPLQELD(1146).

In general, our approach is able to identify potential immunoreactive domains and epitopes of the Spike protein based only on structural and energetic information: our approach correctly predicts between 20 and 80% of the aminoacids engaged by monoclonal Abs in reported X-ray structures. As different (combinations of) antigen residues may be shared among different antibodies in a polyclonal response (such as the one taking place in the host organism), capturing even the minimal sequence endowed with potential

immunoreactivity can aptly represent a useful step towards designing molecules able to elicit Abs capable of interfering with viral entry or replication.

In this framework, sequences predicted to be reactive using the restrictive epitope definition (5% cutoff) can be used for generating optimized antigens in the form of synthetic peptide epitopes. Engineering such epitopes would entail the synthesis of conformationally preorganized peptidomimetics of the 'natural' reactive regions—with intra- and extracellular stability enhanced through, e.g, a combination of natural and non-natural aminoacids—which could reproduce the main structural and energetic conditions required to elicit a humoral immune response, as well as constituting candidates for vaccine development. Furthermore, reactive peptides thus identified may be suitable for use as baits in serologic diagnostic applications (e.g., in ELISA assays and in microarrays), to capture and detect to capture and detect not only circulating antibodies that are expressed in response to SARS-CoV-2 infections but also those that are endowed with neutralization activity and thus potentially predicting the infection outcome. As a further application, these peptide-based baits can represent a useful tool for isolating new mAbs and the screening of small molecules for drug development.

One the most significant aspects of our approach is that the S protein's entire glycan shield is explicitly accounted for in the prediction of the immunoreactive regions. Indeed, the various oligosaccharide chains appear to behave differently (see differential coloring of oligosaccharide chains in Figure 1). In light of their stronger energetic coupling to other areas of the protein, some of the glycans are not recognized as epitopes, and thus form an integral part of the stabilizing intramolecular interaction network of S (white chains in Figure 1B); on

the other hand, MLCE also identifies a second subset of poorly coupled oligosaccharides as potentially reactive epitopes (or part thereof) (colored oligosaccharide chains in Figure 1B; carbohydrate cluster in S2 targeted by the glycan-dependent antibody HIV-1 bnAb 2G12, see Figure 1B, 2), highlighting potential vulnerable spots in the glycan shield that could be exploited to design novel immunoreagents and vaccine candidates.

The portion of the glycan shield falling within the former category thus mainly serves to *protect* the protein from recognition by antibodies and consequently enhances viral infectiousness, as well as providing extra structural support. Two such glycans are further exemplified in **Figure 6**.

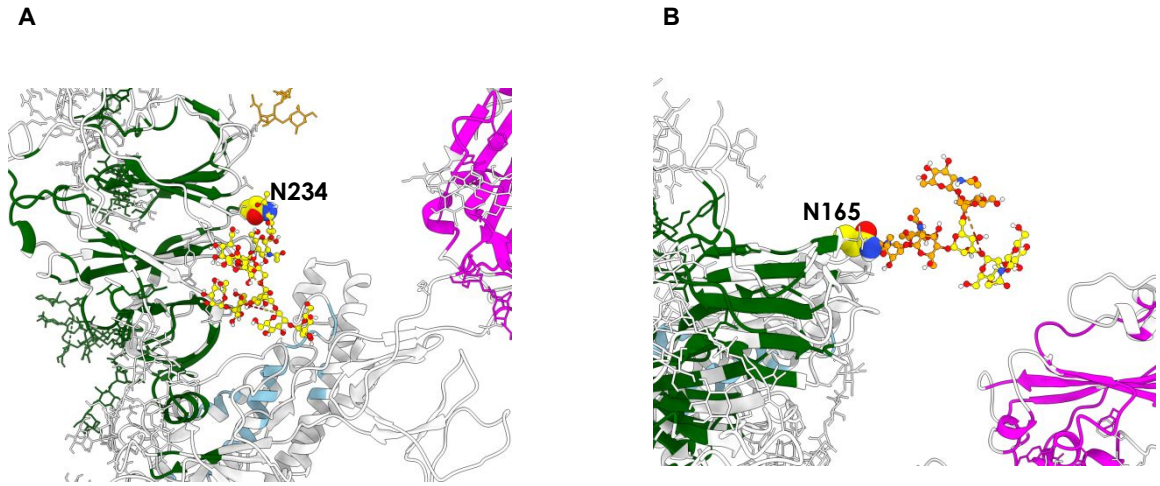


Figure 6. Glycans with different roles on Spike. A) The glycan chain attached to N234, which is predicted to be part of the networks of stabilizing interactions within the protein. **B)** The glycan chain attached to N165 is predicted to play a double role, a stabilizing one (yellow units) and an immunoreactive one (orange units).

The first is the entire oligosaccharide fragment bound to N234 (Figure 6A), which is recognized by Amaro and coworkers as being crucial in 'propping up' the RBD[12]. Experimental deletion of *N*-glycans at this position by way of a mutation to Ala significantly

modifies the conformational landscape of the protein's RBD [53]. The second is the portion of the N165-linked glycan whose subunits are rendered in yellow (**Figure 6B**): consistent with experimental studies indicating N165-linked oligosaccharides as structural modulators [53], we also find that the portion in question is *not* identified as a potential epitope, being consequently involved in diverting antibodies from targeting the region around N165 and thus preserving control of the S protein's structural dynamics.

Reflecting the multifaceted roles of the glycan shield, the remaining part of the N165-linked glycan (**Figure 6B; orange**) appears instead to belong to the category of glycans that *are* potentially able to act as epitopes, since, unlike the part in yellow, we do detect it to be decoupled from the rest of the protomer.

It is particularly significant to underline that MLCE, whose physical basis is to identify non-optimized interaction networks, detects peptidic epitopes even when they are in proximity of (optimized, non-immunogenic) shielding carbohydrates. In light of this, it is reasonable to suggest that the protective effect of these particular carbohydrates may be circumvented and neutralized by exposing the underlying peptidic substructures. Furthermore, information on oligosaccharides identified as epitope constituents can be exploited to design glycomimics or glycosylated peptides as synthetic epitopes.

The latter aspect is indeed particularly relevant: small synthetic molecules that mimic antigenic determinants (and effectively act as their minimal surrogates) offer enticing opportunities to develop immunoreagents with superior characteristics in terms of ease of handling, reproducibility of batch-to-batch production, ease of purification, sustainable cost,

and better stability under a variety of conditions. Furthermore, production of these molecules greatly reduces the risk of cross-reactivity with any copurified antigens, which is instead rife when dealing with recombinant proteins. In contrast to smaller peptides or sugar-decorated peptidomimetics, a full-length recombinant antigenic protein (or any protein-based detection device) would typically require more stringent conditions (e.g., in terms of temperature and humidity) for storage, transport, and management in order to preserve the protein in its properly folded active form. The same would be true for other vaccinal solutions such as deactivated pathogens.

Overall, our work confirms how simple and transparent structural and physico-chemical understanding of the molecule that is the key player in SARS-CoV-2 viral infection can be harnessed to guide the prediction of (in some cases experimentally confirmed) regions, that are involved in immune recognition and to understand its molecular bases. Agreement with experiment confirms that knowledge generated in the process has the potential of being translated into new molecules for vaccine and diagnostic development. In this context, we have also identified potentially reactive regions in the S protein stalk that are currently under experimental synthesis and testing.

Furthermore, potential functional implications offered by the approach are illustrated by the fact that domains/regions relevant for the protein's biological activation are naturally identified. This renders the approach well-suited to identify subtle functional variations in possible mutants of the S proteins that are expected to emerge as a result of further viral diffusion and host adaptation. Finally, the possibility of accurately partitioning such a

complex system in functional subunits could aptly be exploited in the parameterization of coarse-grained models to simulate the system at longer timescales.

This kind of structure-based computational approach can clearly expand the scope of simple structural analysis and molecular simulations. In applicative terms, generation of synthetic libraries based on predicted/identified epitopes (with possible addition of sugars) would definitely boost selection and screening of antigens for vaccine development.

The methodology used in this work is the following:

Coordinates of the fully glycosylated SARS-CoV-2 S protein's 'RBD up' protomer featured in this work originate from molecular dynamics (MD) simulations by Woods and coworkers [13], based on PDB ID: 6VSB. Throughout this work, we retain exactly the same forcefield parameters used by Woods *et al.* in their MD simulations: all residues except glycosylated asparagines are treated using the *ff14SB* forcefield, [54] whereas glycans and glycosylated asparagines are modeled using the *GLYCAM_06j* forcefield [52].

Clustering is based on root-mean-squared deviation of Cα atoms of the RBD domain in the 'RBD up' protomer, and performed with the *cpptraj* utility in *AmberTools* (version 17)[55], after concatenating all six independent MD replicas [13] and aligning them with the 'autoimage' command. The chosen method is the Hierarchical Agglomerative Algorithm [56], with an epsilon value of 0.5. From each of the three most populated clusters, we isolate one representative frame, from which we retain the 'RBD up' protomer and its glycans, whilst again using *cpptraj* to discard all solvent molecules, ions, and the two 'RBD down' protomers. All subsequent calculations on these three 'RBD up' protomer models are listed chronologically in the subsections that follow.

A 200-step minimization of each of the three protomer models is carried out using the default procedure (i.e., steepest descent for 10 steps; then conjugate gradient) implemented in the MD engine *sander* in the *AMBER* software package (version 18))[55]. Protomers are minimized using the generalized Born (GB) implicit solvent model as parametrized by Onufriev et al. [57], with a universal 12.0 Å cutoff applied in the calculation of Lennard-Jones and Coulomb interactions (neither of which are calculated beyond this limit). For this stage,

concentration of (implicit) mobile counterions in the GB model is set to 0.1 M, and the solvent-accessible surface area (SASA) is computed according to the LCPO method (linear combinations of pairwise overlaps) [58].

MM/GBSA calculations [41] are performed on each of the three minimized 'RBD up' protomers using the dedicated mm_pbsa.pl utility in *AmberTools* (version 17). The purpose of these calculations is to obtain a breakdown of nonbonded energy interactions (i.e., electrostatic, van der Waals, implicit solvation contributions and, in this case, 1-4 interactions) between every possible pair of residues in the protomer (aminoacids and monosaccharides alike): for a protomer composed of *N* residues, this leads to a symmetric $N \times N$ interaction matrix M_{ij}. [59-60]

The implicit GB solvation model used in these calculations is identical to the one used in the preceding minimization step (vide supra), except that the implicit ion concentration is set to 0.0 M, and SASA is computed with the ICOSA method (based on icosahedra around each atom that are progressively refined to spheres).

The symmetric interaction matrix M_{ij} obtained from separate MM/GBSA calculations (vide supra) on each of the three S protein protomer models under study (with N residues) can be expressed in terms of its eigenvalues and eigenvectors as

$$M_{ij} = \sum_{\alpha=1}^{N} \lambda_\alpha v_i^\alpha v_j^\alpha$$

where λ_α is the α-th eigenvalue and v_i^α is the i^{th} component of the corresponding eigenvector.

It was previously shown in a number of cases that `eigenvector` (v^1), also called *first eigenvector,* associated with the lowest eigenvalueλ_1 allows to identify most of the crucial aminoacids necessary for the stabilization of a protein fold, and consequently those aminoacids that are minimally coupled to such core. The latter were shown to correspond to potential interaction regions.

In the case of multidomain proteins such as S, the first eigenvector is not sufficient, and more eigenvectors are needed to capture the essential interactions for folding/stability and binding. The interaction matrix M_{ij} is thus decomposed instead via the alternative approach developed by Genoni *et al.* [30]. In this scenario, the aim is to select the smallest set of N_e eigenvectors that cover the largest part of residues (i.e., components) with the minimum redundancy under the assumption that: (a) for each domain there should exist only one associated eigenvector recapitulating its most significant interactions; (b) each "domain eigenvector" has a block structure whereby its significant components correspond to the residues belonging to the identified domain; (c) combination of all significant blocks covers all residues in the protein. Matrix M_{ij} can thus be reformulated as a simplified matrix $\tilde{M}_{ij}$:

$$\tilde{M}_{ij} = \sum_{\alpha=1}^{Ne} \lambda_\alpha v_i^\alpha v_j^\alpha$$

(where this time the sum occurs over N_e essential eigenvectors instead of N residues). As detailed by Genoni *et al.* [30], the essential folding matrix $\tilde{M}_{ij}$is subsequently further filtered through a symbolization process to emphasize the significant non-bonded interaction,

yielding $\tilde{M}_{ij}^{S}\tilde{M}_{ij}^{S}$and finally subjected to a proper clustering procedure leading to domain identification.

The final simplified matrix $\tilde{M}_{ij}^{S}$ resulting from domain decomposition thus only reports those residue pairs in the protomer that exhibit the strongest and weakest energetic interactions.

Final epitope predictions are made using the Matrix of Local Coupling Energies method (MLCE), in which analysis of a given protein's energetic properties is combined with that of its structural determinants [25, 27]. This approach allows to identify nonoptimized, contiguous regions on the protein surface that are deemed to have minimal coupling energies with the rest of the structure, and that have a greater propensity for recognition by Abs or other binding partners.

The MLCE procedure entails cross-comparison of the simplified pairwise residue-residue energy interaction matrix $\tilde{M}_{ij}$ resulting from domain decomposition (vide supra) with a pairwise residue-residue contact matrix C_{ij}. The latter matrix namely considers a pair of residues to be spatially contiguous (i.e., 'in contact') if they are closer than an arbitrary 6.0 Å-threshold; contact distances are measured between Cβ atoms in the case of non-glycine aminoacid residues, H atoms in the case of glycine residues, and between C1 atoms in the case of glycan residues.

The Hadamard product of the two matrices yields the matrix of the local pairwise coupling energies $MLCE_{ij}$:

$$MLCE_{ij} = \tilde{M}_{ij} \cdot C_{ij}$$

Deriving the MLCE matrix allows to rank spatially contiguous residue pairs with respect to the strength of their energetic interactions (weakest to strongest). Selection of proximal pairs

showing the weakest coupling with the rest of the protein ultimately defines putative epitopes; two distinct selections are carried out on the basis of two possible weakness (softness) cutoffs (5% or 15%), corresponding to the top 5% or 15% spatially contiguous residue pairs with the lowest energetic interactions.

Acknowledgement. The authors thank Prof. Robert J. Woods and Dr. Oliver Grant (University of Georgia) for kindly providing atomistic molecular dynamics simulations of the fully glycosylated spike proteins. GC gratefully acknowledges Dr. Riccardo Capelli, Dr. Claudio Peri, and Dr. Guido Scarabelli for previous work on epitope prediction. This research was supported by a Grant from “Cassa di Risparmio di Padova e Rovigo (CARIPARO) PROGETTI DI RICERCA SUL COVID-19” to GC and AR.

Supporting Information

The supporting information is available free of charge at https://pubs.acs.org

- Document with instructions on the structure-prediction files uploaded (PDF);
- Machine-Readable Projection of the MLCE Matrix onto the Spike Protein (ZIP);
- The code to perform the analysis (ZIP). The code is also is available on GitHub at the following address: https://github.com/colombolab/MLCE

References

1. Kupferschmidt, K.; Cohen, J., Race to find COVID-19 treatments accelerates. *Science* **2020,** *367* (6485), 1412.
2. Li, G.; De Clercq, E., Therapeutic options for the 2019 novel coronavirus (2019-nCoV). *Nat. Rev. Drug. Disc.* **2020,** *19*, 149-150.
3. Liu, C.; Zhou, Q.; Li, Y.; Garner, L. V.; Watkins, S. P.; Carter, L. J.; Smoot, J.; Gregg, A. C.; Daniels, A. D.; Jervey, S.; Albaiu, D., Research and Development on Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Diseases. *ACS Central Science* **2020,** *6* (3), 315-331.
4. Romagnoli, S.; Peris, A.; De Gaudio, A. R.; Geppetti, P., SARS-CoV-2 and COVID-19: From the Bench to the Bedside. *Physiological Reviews* **2020,** *100* (4), 1455-1466.
5. Andersen, K. G.; Rambaut, A.; Lipkin, W. I.; Holmes, E. C.; Garry, R. F., The proximal origin of SARS-CoV-2. *Nature Medicine* **2020,** *26* (4), 450-452.
6. Weiss, S. R.; Leibowitz, J. L., Coronavirus pathogenesis. *Adv. Virus Res* **2011,** *81*, 85-164.
7. 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* **2020**, eabb2762.
8. Tortorici, M. A.; Veesler, D., Structural insights into coronavirus entry. *Adv Virus Res* **2019,** *105*, 93-116.
9. Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C.-L.; Abiona, O.; Graham, B. S.; McLellan, J. S., Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **2020**, eabb2507.
10. Walls, A. C.; Tortorici, M. A.; Bosch, B.-J.; Frenz, B.; Rottier, P. J. M.; DiMaio, F.; Rey, F. A.; Veesler, D., Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. *Nature* **2016,** *531* (7592), 114-117.
11. Walls, A. C.; Park, Y.-J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D., Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020,** *181* (2), 281-292.e6.
12. Casalino, L.; Gaieb, Z.; Dommer, A. C.; Harbison, A. M.; Fogarty, C. A.; Barros, E. P.; Taylor, B. C.; Fadda, E.; Amaro, R. E., Shielding and Beyond: The Roles of Glycans in SARS-CoV-2 Spike Protein. *bioRxiv* **2020**, 2020.06.11.146522.

13. Grant, O. C.; Montgomery, D.; Ito, K.; Woods, R. J., 3D Models of glycosylated SARS-CoV-2 spike protein suggest challenges and opportunities for vaccine development. *bioRxiv* **2020**, 2020.04.07.030445.

14. Sikora, M.; von Bülow, S.; Blanc, F. E. C.; Gecht, M.; Covino, R.; Hummer, G., Map of SARS-CoV-2 spike epitopes not shielded by glycans. *bioRxiv* **2020**, 2020.07.03.186825.

15. Zhao, P.; Praissman, J. L.; Grant, O. C.; Cai, Y.; Xiao, T.; Rosenbalm, K. E.; Aoki, K.; Kellman, B. P.; Bridger, R.; Barouch, D. H.; Brindley, M. A.; Lewis, N. E.; Tiemeyer, M.; Chen, B.; Woods, R. J.; Wells, L., Virus-Receptor Interactions of Glycosylated SARS-CoV-2 Spike and Human ACE2 Receptor. *Cell Host & Microbe*.

16. Wang, Y.; Liu, M.; Gao, J., Enhanced receptor binding of SARS-CoV-2 through networks of hydrogen-bonding and hydrophobic interactions. *Proceedings of the National Academy of Sciences* **2020,** *117* (25), 13967.

17. 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. *The Journal of Physical Chemistry Letters* **2020,** *11* (12), 4785-4790.

18. Peri, C.; Gagni, P.; Combi, F.; Gori, A.; Chiari, M.; Longhi, R.; Cretich, M.; Colombo, G., Rational epitope design for protein targeting. . *ACS Chemical Biology* **2013,** *8*, 397-404

19. Gourlay, L.; Peri, C.; Bolognesi, M.; Colombo, G., Structure and Computation in Immunoreagent Design: From Diagnostics to Vaccines. *Trends in Biotechnology* **2017,** *35* (12), 1208-1220.

20. Smith, C. C.; Entwistle, S.; Willis, C.; Vensko, S.; Beck, W.; Garness, J.; Sambade, M.; Routh, E.; Olsen, K.; Kodysh, J.; O'Donnell, T.; Haber, C.; Heiss, K.; Stadler, V.; Garrison, E.; Grant, O. C.; Woods, R. J.; Heise, M.; Vincent, B. G.; Rubinsteyn, A., Landscape and Selection of Vaccine Epitopes in SARS-CoV-2. *bioRxiv : the preprint server for biology* **2020**, 2020.06.04.135004.

21. De Gregorio, E.; Rappuoli, R., From empiricism to rational design: a personal perspective of the evolution of vaccine development. *Nat. Rev. Immunol.* **2014,** *14* (7), 505-514.

22. Rappuoli, R.; Bottomley, M. J.; D'Oro, U.; Finco, O.; De Gregorio, E., Reverse vaccinology 2.0: Human immunology instructs vaccine antigen design. *The Journal of Experimental Medicine* **2016,** *213*, 469-481.

23. Thomas, S.; Dilbarova, R.; Rappuoli, R., Future Challenges for Vaccinologists. *Methods Mol Biol* **2016,** *1403*, 41-55.

24. Grant, O.; Woods, R. J., *Glycosylated Swiss-model molecular dynamics trajectory of SARS-CoV-2 spike glycoprotein*. 2020.

25. Scarabelli, G.; Morra, G.; Colombo, G., Predicting interaction sited from the energetics of isolated proteins: a new approach to epitope mapping. *Biophys. J.* **2010,** *98* (9), 1966-1975.

26. Gourlay, L. J.; Peri, C.; Ferrer-Navarro, M.; Conchillo-Solé, O.; Gori, A.; Rinchai, D.; Thomas, R. J.; Champion, O. L.; Michell, S. L.; Kewcharoenwong, C.; Nithichanon, A.; Lassaux, P.; Perletti, L.; Longhi, R.; Lertmemongkolchai, G.; Titball, R. W.; Daura, X.; Colombo, G.; Bolognesi, M., Exploiting the Burkholderia pseudomallei Acute Phase Antigen BPSL2765 for Structure-Based Epitope Discovery/Design in Structural Vaccinology. *Chem. Biol.* **2013,** *20*, 1147-1156.

27. Marchetti, F.; Capelli, R.; Rizzato, F.; Laio, A.; Colombo, G., The Subtle Trade-Off between Evolutionary and Energetic Constraints in Protein-Protein Interactions. *J. Phys. Chem. Lett.* **2019,** *10* (7), 1489-1497.

28. Paladino, A.; Woodford, M. R.; Backe, S. J.; Sager, R. A.; Kancherla, P.; Daneshvar, M. A.; Chen, V. Z.; Bourboulia, D.; Ahanin, E. F.; Prodromou, C.; Bergamaschi, G.; Strada, A.; Cretich, M.; Gori, A.; Veronesi, M.; Bandiera, T.; Vanna, R.; Bratslavsky, G.; Serapian, S. A.; Mollapour, M.; Colombo, G., Chemical Perturbation of Oncogenic Protein Folding: from the Prediction of Locally Unstable Structures to the Design of Disruptors of Hsp90–Client Interactions. *Chemistry – A European Journal* **2020,** *26* (43), 9459-9465.

29. Serapian, S. A.; Colombo, G., Designing Molecular Spanners to Throw in the Protein Networks. *Chemistry – A European Journal* **2020,** *26* (21), 4656-4670.

30. Genoni, A.; Morra, G.; Colombo, G., Identification of Domains in Protein Structures from the Analysis of Intramolecular Interactions. *J. Phys. Chem. B.* **2012,** *116* (10), 3331-3343.

31. Soriani, M.; Petit, P.; Grifantini, R.; Petracca, R.; Gancitano, G.; Frigimelica, E.; Nardelli, F.; Garcia, C.; Spinelli, S.; Scarabelli, G.; Fiorucci, S.; Affentranger, R.; Ferrer-Navarro, M.; Zacharias, M.; Colombo, G.; Vuillard, L.; Daura, X.; Grandi, G., Exploiting antigenic diversity for vaccine design: the chlamydia ArtJ paradigm. *J. Biol. Chem.* **2010,** *285* (39), 30126-30138.

32. Lassaux, P.; Peri, C.; Ferrer-Navarro, M.; Gourlay, L.; Gori, A.; Conchillo-Solé, O.; Rinchai, D.; Lertmemongkolchai, G.; Longhi, R.; Daura, X.; Colombo, G.; Bolognesi, M., A

structure-based strategy for epitope discovery in Burkholderia pseudomallei OppA antigen. *Structure* **2013,** *21*, 1-9.

33. Gourlay, L. J.; Lassaux, P.; Thomas, R. J.; Peri, C.; Conchillo-Sole, O.; Nithichanon, A.; Ferrer-Navarro, M.; Vila, J.; Daura, X.; Lertmemongkolchai, G.; Titball, R.; Colombo, G.; Bolognesi, M., Flagellar subunits as targets for structure-based epitope discovery approaches and melioidosis vaccine development. *Febs Journal* **2015,** *282*, 338-338.

34. Gourlay, L. J.; Thomas, R. J.; Peri, C.; Conchillo-Sole, O.; Ferrer-Navarro, M.; Nithichanon, A.; Vila, J.; Daura, X.; Lertmemongkolchai, G.; Titball, R.; Colombo, G.; Bolognesi, M., From crystal structure to in silico epitope discovery in the Burkholderia pseudomallei flagellar hook-associated protein FlgK. *Febs Journal* **2015,** *282* (7), 1319-1333.

35. Nithichanon, A.; Rinchai, D.; Gori, A.; Lassaux, P.; Peri, C.; Conchillio-Sole, O.; Ferrer-Navarro, M.; Gourlay, L. J.; Nardini, M.; Vila, J.; Daura, X.; Colombo, G.; Bolognesi, M.; Lertmemonkolchai, G., Sequence- and Structure-Based Immunoreactive Epitope Discovery for Burkholderia pseudomallei Flagellin. *Plos Neglected Tropical Diseases* **2015,** *9* (7).

36. Gori, A.; Peri, C.; Quilici, G.; Nithichanon, A.; Gaudesi, D.; Longhi, R.; Gourlay, L.; Bolognesi, M.; Lertmemongkolchai, G.; Musco, G.; Colombo, G., Flexible vs Rigid Epitope Conformations for Diagnostic- and Vaccine-Oriented Applications: Novel Insights from the Burkholderia pseudomallei BPSL2765 Pa13 Epitope. *Acs Infectious Diseases* **2016,** *2* (3), 221-230.

37. Gori, A.; Sola, L.; Gagni, P.; Bruni, G.; Liprino, M.; Peri, C.; Colombo, G.; Cretich, M.; Chiari, M., Screening Complex Biological Samples with Peptide Microarrays: The Favorable Impact of Probe Orientation via Chemoselective Immobilization Strategies on Clickable Polymeric Coatings. *Bioconjugate Chemistry* **2016,** *27* (11), 2669-2677.

38. Sola, L.; Gagni, P.; D'Annessa, I.; Capelli, R.; Bertino, C.; Romanato, A.; Damin, F.; Bergamaschi, G.; Marchisio, E.; Cuzzocrea, A.; Bombaci, M.; Grifantini, R.; Chiari, M.; Colombo, G.; Gori, A.; Cretich, M., Enhancing Antibody Serodiagnosis Using a Controlled Peptide Coimmobilization Strategy. *ACS Infectious Diseases* **2018,** *4* (6), 998-1006.

39. Bergamaschi, G.; Fassi, E. M. A.; Romanato, A.; D'Annessa, I.; Odinolfi, M. T.; Brambilla, D.; Damin, F.; Chiari, M.; Gori, A.; Colombo, G.; Cretich, M., Computational Analysis of Dengue Virus Envelope Protein (E) Reveals an Epitope with Flavivirus Immunodiagnostic Potential in Peptide Microarrays. *International journal of molecular sciences* **2019,** *20* (8), 1921.

40. Grant, O. C.; Montgomery, D.; Ito, K.; Woods, R. J., Analysis of the SARS-CoV-2 spike protein glycan shield reveals implications for immune recognition. *Sci. Rep.* **2020,** *In press*.

41. Genheden, S.; Ryde, U., The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov.* **2015,** *10* (5), 449-461.

42. Yuan, M.; Liu, H.; Wu, N. C.; Lee, C.-C. D.; Zhu, X.; Zhao, F.; Huang, D.; Yu, W.; Hua, Y.; Tien, H.; Rogers, T. F.; Landais, E.; Sok, D.; Jardine, J. G.; Burton, D. R.; Wilson, I. A., Structural basis of a shared antibody response to SARS-CoV-2. *Science* **2020**, eabd2321.

43. Chi, X.; Yan, R.; Zhang, J.; Zhang, G.; Zhang, Y.; Hao, M.; Zhang, Z.; Fan, P.; Dong, Y.; Yang, Y.; Chen, Z.; Guo, Y.; Zhang, J.; Li, Y.; Song, X.; Chen, Y.; Xia, L.; Fu, L.; Hou, L.; Xu, J.; Yu, C.; Li, J.; Zhou, Q.; Chen, W., A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science* **2020**, eabc6952.

44. Barnes, C. O.; West, A. P., Jr.; Huey-Tubman, K. E.; Hoffmann, M. A. G.; Sharaf, N. G.; Hoffman, P. R.; Koranda, N.; Gristick, H. B.; Gaebler, C.; Muecksch, F.; Lorenzi, J. C. C.; Finkin, S.; Hägglöf, T.; Hurley, A.; Millard, K. G.; Weisblum, Y.; Schmidt, F.; Hatziioannou, T.; Bieniasz, P. D.; Caskey, M.; Robbiani, D. F.; Nussenzweig, M. C.; Bjorkman, P. J., Structures of Human Antibodies Bound to SARS-CoV-2 Spike Reveal Common Epitopes and Recurrent Features of Antibodies. *Cell* **2020**, S0092-8674(20)30757-1.

45. Zhou, H.; Chen, Y.; Zhang, S.; Niu, P.; Qin, K.; Jia, W.; Huang, B.; Zhang, S.; Lan, J.; Zhang, L.; Tan, W.; Wang, X., Structural definition of a neutralization epitope on the N-terminal domain of MERS-CoV spike glycoprotein. *Nature Communications* **2019,** *10* (1), 3068.

46. Zheng, Z.; Monteil, V. M.; Maurer-Stroh, S.; Yew, C. W.; Leong, C.; Mohd-Ismail, N. K.; Arularasu, S. C.; Chow, V. T. K.; Pin, R. L. T.; Mirazimi, A.; Hong, W.; Tan, Y.-J., Monoclonal antibodies for the S2 subunit of spike of SARS-CoV cross-react with the newly-emerged SARS-CoV-2. *bioRxiv* **2020**, 2020.03.06.980037.

47. Acharya, P.; Williams, W.; Henderson, R.; Janowska, K.; Manne, K.; Parks, R.; Deyton, M.; Sprenz, J.; Stalls, V.; Kopp, M.; Mansouri, K.; Edwards, R. J.; Meyerhoff, R. R.; Oguin, T.; Sempowski, G.; Saunders, K.; Haynes, B. F., A glycan cluster on the SARS-CoV-2 spike ectodomain is recognized by Fab-dimerized glycan-reactive antibodies. *bioRxiv : the preprint server for biology* **2020**, 2020.06.30.178897.

48. Tang, T.; Bidon, M.; Jaimes, J. A.; Whittaker, G. R.; Daniel, S., Coronavirus membrane fusion mechanism offers a potential target for antiviral development. *Antiviral Research* **2020,** *178*, 104792.

49. Baum, A.; Fulton, B. O.; Wloga, E.; Copin, R.; Pascal, K. E.; Russo, V.; Giordano, S.; Lanza, K.; Negron, N.; Ni, M.; Wei, Y.; Atwal, G. S.; Murphy, A. J.; Stahl, N.; Yancopoulos, G. D.; Kyratsous, C. A., Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. *Science (New York, N.Y.)* **2020,** *369* (6506), 1014-1018.

50. Pinto, D.; Park, Y.-J.; Beltramello, M.; Walls, A. C.; Tortorici, M. A.; Bianchi, S.; Jaconi, S.; Culap, K.; Zatta, F.; De Marco, A.; Peter, A.; Guarino, B.; Spreafico, R.; Cameroni, E.; Case, J. B.; Chen, R. E.; Havenar-Daughton, C.; Snell, G.; Telenti, A.; Virgin, H. W.; Lanzavecchia, A.; Diamond, M. S.; Fink, K.; Veesler, D.; Corti, D., Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature* **2020,** *583* (7815), 290-295.

51. Cao, Y.; Su, B.; Guo, X.; Sun, W.; Deng, Y.; Bao, L.; Zhu, Q.; Zhang, X.; Zheng, Y.; Geng, C.; Chai, X.; He, R.; Li, X.; Lv, Q.; Zhu, H.; Deng, W.; Xu, Y.; Wang, Y.; Qiao, L.; Tan, Y.; Song, L.; Wang, G.; Du, X.; Gao, N.; Liu, J.; Xiao, J.; Su, X.-d.; Du, Z.; Feng, Y.; Qin, C.; Qin, C.; Jin, R.; Xie, X. S., Potent Neutralizing Antibodies against SARS-CoV-2 Identified by High-Throughput Single-Cell Sequencing of Convalescent Patients' B Cells. *Cell* **2020,** *182* (1), 73-84.e16.

52. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J., GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *Journal of Computational Chemistry* **2008,** *29* (4), 622-655.

53. Henderson, R.; Edwards, R. J.; Mansouri, K.; Janowska, K.; Stalls, V.; Kopp, M.; Haynes, B. F.; Acharya, P., Glycans on the SARS-CoV-2 Spike Control the Receptor Binding Domain Conformation. *bioRxiv* **2020**, 2020.06.26.173765.

54. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **2015,** *11* (8), 3696-3713.

55. Case, D. A.; Cheatham Iii, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz Jr, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *Journal of Computational Chemistry* **2005,** *26* (16), 1668-1688.

56. Defays, D., An efficient algorithm for a complete link method. *The Computer Journal* **1977,** *20* (4), 364-366.

57. Onufriev, A.; Bashford, D.; Case, D. A., Modification of the Generalized Born Model Suitable for Macromolecules. *J. Phys. Chem. B.* **2000,** *104*, 3712-3720.

58. Weiser, J.; Shenkin, P. S.; Still, W. C., Approximate Atomic Surfaces from Linear Combinations of Pairwise Overlaps (LCPO). *J. Comput. Chem.* **1999,** *20*, 217-230.

59. Tiana, G.; Simona, F.; De Mori, G. M. S.; Broglia, R. A.; Colombo, G., Understanding the determinants of stability and folding of small globular proteins from their energetics. *Protein Science* **2004,** *13* (1), 113-124.

60. Morra, G.; Colombo, G., Relationship between energy distribution and fold stability: Insights from molecular dynamics simulations of native and mutant proteins. *Proteins: Struct. Funct. and Bioinf.* **2008,** *72* (2), 660-672.