A possible strategy to fight COVID-19: Interfering with spike glycoprotein trimerization

Pietro Bongini [a, c], Alfonso Trezza [b], Monica Bianchini [a], Ottavia Spiga [b], Neri Niccolai [b, d, *]

[a] Department of Information Engineering and Mathematics, University of Siena, Siena, 53100, Italy
[b] Department of Biotechnology, Chemistry and Pharmacy, University of Siena, Siena, 53100, Italy
[c] Department of Information Engineering, University of Florence, Florence, 50139, Italy
[d] Le Ricerche Del BarLume Free Association, Monteroni D'Arbia, 53014, Italy

ARTICLE INFO

ABSTRACT

The recent release of COVID-19 spike glycoprotein allows detailed analysis of the structural features that are required for stabilizing the infective form of its quaternary assembly. Trying to disassemble the trimeric structure of COVID-19 spike glycoprotein, we analyzed single protomer surfaces searching for concave moieties that are located at the three protomer-protomer interfaces. The presence of some druggable pockets at these interfaces suggested that some of the available drugs in Drug Bank could destabilize the quaternary spike glycoprotein formation by binding to these pockets, therefore interfering with COVID-19 life cycle. The approach we propose here can be an additional strategy to fight against the deadly virus. Ligands of COVID-19 spike glycoprotein that we have predicted in the present computational investigation, might be the basis for new experimental studies *in vitro* and *in vivo*.

© 2020 Published by Elsevier Inc.

1. Introduction

Nowadays COVID-19 infection represents a terrible and unexpected threat to human life. It is rather obvious that, waiting for efficient vaccines, antiviral drugs and, eventually, herd immunity, any attempt to fight against the deadly virus needs to be performed. Available tools of structural and computational Biology are particularly suited to obtain predictions for new therapeutic strategies through fast and safe procedures if we have a reliable structural basis. The fact that the quaternary structure of COVID-19 spike (S) glycoprotein has been resolved [1] and is now freely downloadable from the Protein Data Bank (PDB) opens to Structural Biology new perspectives to stop the viral infection. The S glycoprotein, indeed, is central for COVID-19 infection as it mediates attachment of virions to the host cell receptor, it is involved in cell-to-cell fusion and induces neutralizing antibodies, bearing also virulence determinants [2].

Assembly of protomers into the bioactive form of S glycoprotein

trimeric structure has been experimentally proved to be the rate-limiting step for transmissible gastroenteritis coronavirus (TGEV) infecting cycle [3]. The close similarity between S glycoprotein of TGEV and COVID-19, as controlled in the present study, suggests that the same hindrances that *in vitro* occur for TGEV S glycoprotein assembly [3] should also hold in the COVID-19 case. Thus, being S glycoprotein trimerization probably not a so direct and fast process, we might have the chance to interfere with the quaternary structure assembly that COVID-19 needs for infecting human cells.

Even though protein-protein adducts are in general characterized by flat interacting surfaces [4], we have explored whether convex moieties that can be suitable for small molecule binding are present at protomer-protomer interfaces of S glycoprotein. In the present report, we describe our search for potential binding sites in S glycoprotein protomers and their corresponding possible ligands.

2. Methods

From the PDB we have retrieved the quaternary structure of COVID-19 S glycoprotein, PDB ID 6VSB [1], as the basis for the present analysis. After removing all heteroatoms from 6VSB PDB file, as they would interfere with our calculations, we defined atom depths and pocket locations for each of the three COVID-19 S glycoprotein protomers. Atom depths were obtained by using

* Corresponding author. Department of Biotechnology, Chemistry and Pharmacy, University of Siena, Siena, 53100, Italy.

E-mail addresses: pietro.bongini@gmail.com (P. Bongini), trezzaalfonso@gmail.com (A. Trezza), monica@diism.unisi.it (M. Bianchini), ottavia.spiga@unisi.it (O. Spiga), neri.niccolai@unisi.it (N. Niccolai).
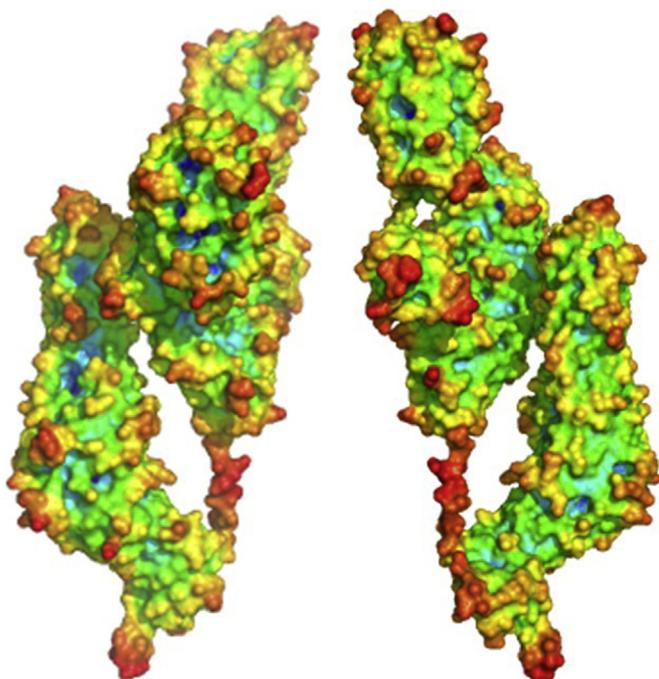
Fig. 1. Front-rear view of protomer A from COVID-19 S glycoprotein: structure is obtained from PDB file with ID code 6VSB; protomer surface is colored according to atom depths.

Table 1
Structural features of filtered surface pockets.

#	interface[a]	residues[b]	DS[c]	Volume[d] ($Å^3$)
1	A_C (85.7)	14	0.94	311.7
2	A_B (83.3)	18	0.90	493.8
3	B_A (93.7)	16	0.89	409.8
4	B_C (77.7)	18	0.90	557.6
5	B_A (92.8)	14	0.91	339.6
6	C_A (80.9)	21	0.97	590.1

[a] Protomer pocket position; index refers to the other protomer involved in the interface; in parenthesis percent overlapping area between a pocket and PISA-defined protomer interface.
[b] Number of residues forming the pocket (details are given in Table 2).
[c] Druggability Scores from PockDrug-Server.
[d] Volumes calculated by $EPOS^{BP}$.

SADIC algorithm [5] through the on-line service [6]. PISA analysis [7] yielded the complete list of residues located at the S glycoprotein protomer interfaces. Pockets located at the protomer surface were identified by using $EPOS^{BP}$ [8]. Autodock VinaXB [9] was used to calculate free energy of interaction between COVID-19 S glycoprotein at pocket #1 and #6 with ligands contained in the Drug-Bank database v.5.1.5 [10]. Ligad toxicity has been analyzed according to ProTox II [11]. Open source PyMOL v. 1.7.1.0 has been used for structural data analysis and presentation.

3. Results and discussion

By comparing S glycoprotein sequences of COVID-19 and TGEV coronaviruses, we gave a preliminary check on the central issue of this investigation, *i.e.* the possibility that the *in vitro* observed problematic assembly of TGEV quaternary S glycoprotein structure [3] would hold also for the one of COVID-19. On March 28 of the current year, NCBI GenBank [12] reported 99 S glycoprotein sequences for COVID-19. One of these sequences, YP_009724390.1, and one representative of TGEV, AAY22404.1, were aligned by using the EMBL-EBI tool Pairwise Sequence Alignment [13]. In the latter alignment the S glycoprotein S2 domain, the most critical region to stabilize the trimeric form of the S glycoprotein, showed 81% sequence similarity, ensuring the reliability of our initial assumption.

Then, from the PDB we have retrieved the quaternary structure of COVID-19 S glycoprotein. Before analyzing the complex pattern of S glycoprotein trimerization, we have depleted the PDB structure from all heteroatoms to simplify our atom depth calculations. Thus, from a molecular visualization accounting for individual atom depth [5], in each of the three S glycoprotein protomers, we could easily identify many surface pockets, see Fig. 1. It is apparent that the surface of S glycoprotein protomer A, in the PDB nomenclature, has plenty of deep concave regions. Afterward, we have carried out a quantitative analysis on each of the three surfaces of S glycoprotein protomers by using $EPOS^{BP}$ software [8] finding 143 pockets with volumes ranging from 190 to 850 $Å^3$. Then, we have characterized the amino acid content of the three protomer-protomer interfaces by using the EBI PISA tool [7]. Overlapping of $EPOS^{BP}$ surface pockets with PISA defined interfaces yielded initial clues of COVID-19 S glycoprotein moieties that could be the target of our study. All the pockets located at protomer-protomer interfaces were screened for druggability according to PockDrug-Server [14].

Hence, to restrict our investigation to the most reliable predictions of potential ligand binding sites on COVID-19 protomer surfaces, we have filtered all $EPOS^{BP}$ defined pockets to consider only the ones fulfilling these limitations: a) overlapping with interface surfaces higher than 70%, b) pocket volume >300 $Å^3$, c) druggability scores >0.7, d) located in the S2 domain of S glycoprotein region and e) the number of pocket residues >10. It must be underlined that condition d is particularly important, as stem moieties of S glycoprotein seem to be the most critical ones for achieving quaternary assembly.

This procedure restricted all $EPOS^{BP}$ defined pockets to 6 cases. Structural features of these pockets, well representing the three protomer-protomer S glycoprotein interfaces, are summarized in Tables 1 and 2. The six pockets reported in Tables 1 and 2, fulfilling all our limitations, can be considered as good starting points for further medicinal chemistry investigations, all of them exhibiting suitable features for small molecule binding. Data reported in Tables 1 and 2 indicate that, apart from pocket #5, all the other ones are present in two or three protomers, even though with slightly different local structures accounting for their volumes and druggability. This finding proves that local rearrangements of COVID-19 S

Table 2
Amino acid composition of pocket surfaces.

#[a]	Interface residues[b]
1 and 3	Pro792 Ile794 Lys795 Asp796 Phe797 Gly798 Ile882 Thr883 Gln895 Ile 896 Pro897 Phe898 Ala899 Met900 Tyr917 Gln920
2, 4 and 6	Tyr707 Ser708 Asn709 Asn710 Ser711 Ile712 Thr1077 Ala1078 Pro1079 Ala1080 Phe1089 Pro1090 Gly1093 Val1094 Phe1095 Arg1107 Phe1121 Ile1130 Gly1131 Ile1132 Val1133
5	Ala783 Gln784 Val785 Lys786 Gln787 Tyr789 Phe888 Gly889 Ala890 Gly891 Ala892 Ala893 Ser1030 Leu1034

[a] Pocket number from Table 1.
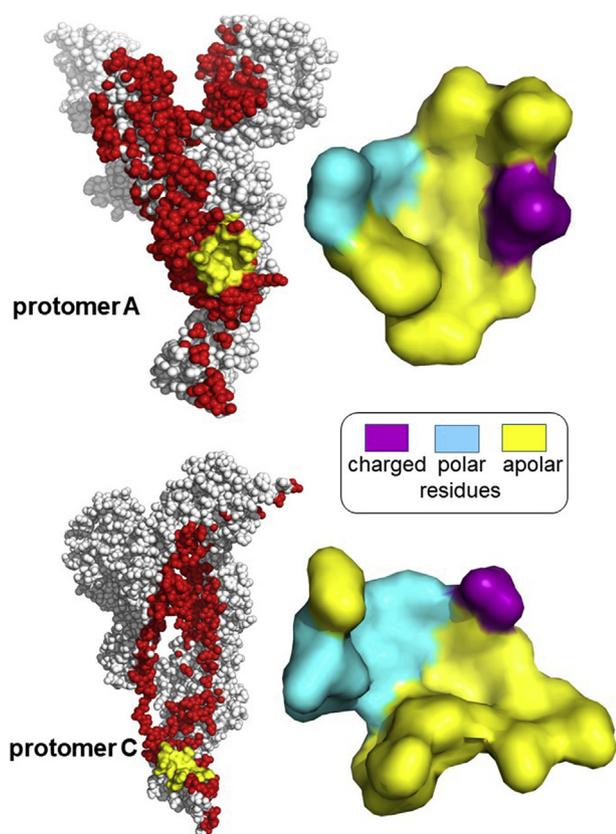[b] Amino acids contributing to surface pocket formation.

Fig. 2. Grey sphere representation of protomer A and C: red colored atoms belong to the interface with protomer C and A respectively. Pocket #1 in protomer A and #6 in protomer C are highlighted with yellow surfaces. Enlarged view of both pockets is shown on the right. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Fig. 3. Protomer A of COVID-19 S glycoprotein bound to ergoloid: detailed view of pocket#1, colored in cyan, bound to ergoloid shown in ball and stick CPK representation. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

glycoprotein are required during quaternary assembling. Fig. 2 summarizes the structural features of pockets #1 and #6.

Before selecting possible ligands that can bind to the above-described protomer pockets, it is mandatory to check the sequence conservation of residues listed in Table 2. Hence, all the 99 S glycoprotein sequences stored in NCBI GenBank [12] were aligned by using Clustal Omega [13]. All residues belonging to COVID-19 S glycoprotein S2 domain, encompassing 660—1273 sequence fragment, exhibited full identity, indicating the fundamental role of this glycoprotein moiety for quaternary structure stability, see Supplementary Material. This finding also suggests that COVID-19 evolution should not influence possible ligand binding to the pockets that we have delineated. We extended our
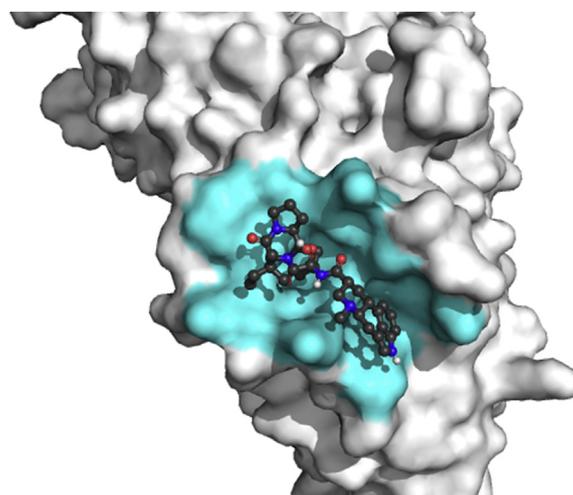
sequence conservation analysis on S glycoprotein not only for COVID-19, but also for similar coronaviruses by using BLAST software [15] and multiple alignment Cobalt tool [16]. As reported in Supplementary Material, alignment of 160 S glycoprotein sequences from NCBI GenBank for a large variety of bat and SARS coronaviruses shows full conservation, apart from Pro792 and Ile882, for all the residues that form pockets reported in Table 1. The latter result adds additional value to the therapeutic use of ligands whose binding to COVID-19 S glycoprotein may interfere with its quaternary structure assembly.

Then, we have looked for the existence of small molecules that can interfere with COVID-19 S glycoprotein trimerization through binding to the pockets that we have delineated on the protomer surfaces. This task was achieved by downloading the content of DrugBank 5.1,5 [10] to obtain two distinct data sets of possible COVID-19 ligands, *i.e.* one set containing 108 nutraceutical drugs and the other set containing 1223 FDA-approved small molecule drugs. Afterward, we carried out a docking simulation between all the small molecules of the two data sets and pockets #1 and #6, among all the ones that are listed in Table 1, for the sake of computing time. By using Autodock VinaXB [9] we have analyzed all the steric and energetic features characterizing the interactions of small molecules in our two data sets. By restricting the docking simulation only to the protomer surface regions where pocket #1 and #6 are located, we obtained a large array of possible ligands.

Table 3
Predicted ligands of COVID-19 protomer surface pockets.

Predicted ligand	Data set[a]	Pocket #[b]	ΔG$_{int}$ (Kcal/mol)[c]	Toxicity[d]
Ergoloid	A	1	−8.0	4
Darifenacin	A	1	−7.7	4
5-methyltetrahydrofolic acid	N	1	−7.7	4
Buclizine	A	1	−7.6	5
Saquinavir	A	1	−7.5	4
Solifenacin	A	1	−7.5	4
Sorafenib	A	1	−7.4	4
Tetrahydrofolic acid	N	1	−7.4	4

[a] N in the case of nutraceutical drugs and A in the case of FDA-approved drugs.
[b] Pocket label from Table 1.
[c] Binding free energy.
[d] Compound toxicity (1 - > high toxicity; 6 - > no toxicity) [11].

Based on their calculated free energy of interaction, we have selected, and listed in Table 3, eight of the best candidates for further experimental investigations.

It is interesting to note that the eight best ligand candidates reported in Table 3 are all binding pocket #1. Fig. 3 reports the sterical features of the interaction between ergoloid and pocket #1.

We believe that the present analysis of protomer-protomer interfaces of COVID-19 S glycoprotein can be a useful starting point for predicting ligands that are already in use for other pathologies and, by interfering with quaternary structure assembly of COVID-19 S glycoprotein, can exhibit therapeutic activity against viral life cycle.

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.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbrc.2020.04.007.

References

[1] D. Wrapp, N. Wang, K.S. Corbett, J.A. Goldsmith, C.L. Hsieh, O. Abiona, B.S. Graham, J.S. McLellan, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation, Science 367 (6483) (2020) 1260–1263, https://doi.org/10.1126/science.abb2507.

[2] W. Spaan, D. Cavanagh, M.C. Horzinek, Coronaviruses: structure and genome expression, J. Gen. Virol. 69 (12) (1988) 2939–2952. PubMed PMID: 3058868.

[3] B. Delmas, H. Laude, Assembly of coronavirus spike protein into trimers and its role in epitope expression, J. Virol. 64 (11) (1990) 5367–5375. PubMed PMID: 2170676.

[4] J.A. Wells, C.L. McClendon, Reaching for high-hanging fruit in drug discovery at protein-protein interfaces, Nature 450 (7172) (2007) 1001–1009. PubMed PMID: 18075579.

[5] D. Varrazzo, A. Bernini, O. Spiga, A. Ciutti, S. Chiellini, V. Venditti, L. Bracci, N. Niccolai, Three-dimensional computation of atom depth in complex molecular structures, Bioinformatics 21 (12) (2005) 2856–2860. PubMed PMID: 15827080.

[6] Simple atom depth index calculator (SADIC), in: http://www.sbl.unisi.it/prococoa/.

[7] E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state, J. Mol. Biol. 372 (3) (2007) 774–797. PubMed PMID: 17681537.

[8] S. Eyrisch, V. Helms, Transient pockets on protein surfaces involved in protein-protein interaction, J. Med. Chem. 50 (15) (2007) 3457–3464. PMID: 17602601.

[9] M.R. Koebel, G. Schmadeke, R.G. Posner, S. Sirimulla, AutoDock VinaXB: implementation of XBSF, new empirical halogen bond scoring function, into AutoDock Vina, J. Cheminform. eCollection (2016), https://doi.org/10.1186/s13321-016-0139-1, 2016.

[10] D.S. Wishart, Y.D. Feunang, A.C. Guo, E.J. Lo, A. Marcu, J.R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox, M. Wilson, DrugBank 5.0: a major update to the DrugBank database for 2018, Nucleic Acids Res. (2017), https://doi.org/10.1093/nar/gkx1037. PubMed: 29126136.

[11] P. Banerjee, A.O. Eckert, A.K. Schrey, R. Preissner, ProTox-II: a webserver for the prediction of toxicity of chemicals, Nucleic Acids Res. 50 (W1) (2018) W257–W263, https://doi.org/10.1093/nar/gky318. PubMed PMID: 29718510.

[12] SARS-CoV-2 (Severe acute respiratory syndrome coronavirus 2) Sequences. https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/.

[13] L. Mullan, Pairwise sequence alignment–it's all about us!, Brief Bioinform. 7 (1) (2006) 113–115, https://doi.org/10.1093/bib/bbk008.

[14] H.A. Hussein, A. Borrel, C. Geneix, M. Petitjean, L. Regad, A.C. Camproux, PockDrug-Server: a new web server for predicting pocket druggability on holo and apo proteins, Nucleic Acids Res. 43 (1) (2015) 436–442, https://doi.org/10.1093/nar/gkv462. PMID: 25956651.

[15] G.M. Boratyn, C. Camacho, P.S. Cooper, G. Coulouris, A. Fong, N. Ma, T.L. Madden, W.T. Matten, S.D. McGinnis, Y. Merezhuk, Y. Raytselis, E.W. Sayers, T. Tao, J. Ye, I. Zaretskaya, BLAST: a more efficient report with usability improvements, Nucleic Acids Res. 41 (2013) 29–33.

[16] J.S. Papadopoulos, R. Agarwala, COBALT: constraint-based alignment tool for multiple protein sequences, Bioinformatics 23 (2007) 1073–1079.