



Unlocking COVID therapeutic targets: A structure-based rationale against SARS-CoV-2, SARS-CoV and MERS-CoV Spike

João Trigueiro-Louro^{a,b,*}, Vanessa Correia^a, Inês Figueiredo-Nunes^b, Marta Gíria^b, Helena Rebelo-de-Andrade^{a,b,*}

^a Antiviral Resistance Lab, Research & Development Unit, Infectious Diseases Department, Instituto Nacional de Saúde Doutor Ricardo Jorge, IP, Av. Padre Cruz, 1649-016 Lisbon, Portugal

^b Host-Pathogen Interaction Unit, Research Institute for Medicines (iMed.Ulisboa), Faculty of Pharmacy, Universidade de Lisboa, Av. Professor Gama Pinto, 1649-003 Lisbon, Portugal



ARTICLE INFO

Article history:

Received 4 June 2020

Received in revised form 20 July 2020

Accepted 22 July 2020

Available online 31 July 2020

Keywords:

Betacoronavirus

Coronavirus disease

Druggability prediction

Novel antiviral targets

SARS-CoV-2

Sequence conservation

Spike protein

ABSTRACT

There are no approved target therapeutics against SARS-CoV-2 or other beta-CoVs. The beta-CoV Spike protein is a promising target considering the critical role in viral infection and pathogenesis and its surface exposed features.

We performed a structure-based strategy targeting highly conserved druggable regions resulting from a comprehensive large-scale sequence analysis and structural characterization of Spike domains across SARSr- and MERSr-CoVs.

We have disclosed 28 main consensus druggable pockets within the Spike. The RBD and SD1 (S1 subunit); and the CR, HR1 and CH (S2 subunit) represent the most promising conserved druggable regions. Additionally, we have identified 181 new potential hot spot residues for the hSARSr-CoVs and 72 new hot spot residues for the SARSr- and MERSr-CoVs, which have not been described before in the literature. These sites/residues exhibit advantageous structural features for targeted molecular and pharmacological modulation.

This study establishes the Spike as a promising anti-CoV target using an approach with a potential higher resilience to resistance development and directed to a broad spectrum of Beta-CoVs, including the new SARS-CoV-2 responsible for COVID-19. This research also provides a structure-based rationale for the design and discovery of chemical inhibitors, antibodies or other therapeutic modalities successfully targeting the Beta-CoV Spike protein.

© 2020 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 license (<http://creativecommons.org/licenses/by/4.0/>).

Abbreviations: aa, amino acid; ACE2, angiotensin-converting enzyme2; Bat-SL-CoVs, bat SARS-like coronavirus; Beta-CoVs, betacoronavirus; CC, conserved cluster; CD, connector domain; CDR, consensus druggable residue; CDP, consensus druggable pocket; CH, central helix; CoVs, coronavirus; CP, cytoplasmic domain; CR, connecting region; CS, conservation score; DGSS, DoGSiteScorer; DPP4, dipeptidyl peptidase-4; FP, fusion peptide; HR1, heptad repeat 1; HR2, heptad repeat 2; hSARSr-CoVs, human Severe acute respiratory syndrome-related coronavirus; MERS-CoVs, middle east respiratory syndrome coronavirus; MERSr-CoVs, middle east respiratory syndrome-related coronavirus; MSA, multiple sequence alignment; NTD, N-terminal domain; nts, nucleotides; PDB, Protein Data Bank; PDS, PockDrug-Server; RBD, Receptor-Binding Domain; S, Spike; SARS-CoVs, severe acute respiratory syndrome coronavirus; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; SARSr-CoVs, severe acute respiratory syndrome-related coronavirus; SD1, subdomain 1; SD2, subdomain 2; Sv, shorter variant; SF, SiteFinder from MOE; SP, small pocket; TMPRSS2, transmembrane protease serine 2; T-RHS, top-ranked hot spots.

* Corresponding authors at: Antiviral Resistance Lab, Research & Development Unit, Infectious Diseases Department, Instituto Nacional de Saúde Doutor Ricardo Jorge, IP, Av. Padre Cruz, 1649-016 Lisbon, Portugal.

E-mail addresses: joao.louro@insa.min-saude.pt (J. Trigueiro-Louro), H.Rebelo.Andrade@insa.min-saude.pt (H. Rebelo-de-Andrade).

1. Introduction

There are no antiviral drugs approved specifically for the prevention or treatment of coronavirus disease 2019 (COVID-19) caused by SARS-CoV-2; or against any other coronavirus (CoVs).

The current recommendations for the clinical management of COVID-19 include infection prevention and control measures and supportive care [1]. Because there is no sufficient information to recommend for or against broad spectrum antivirals, repurposing drugs or immunomodulatory therapy, moderate to critically ill patients have been managed with these alternative therapies through Emergency Use Authorizations, Emergency Investigational New Drug applications, compassionate use or expanded access programs with drug manufacturers [1,2].

The molecular epidemiology of SARS-CoV-2 in the post-pandemic period is uncertain. It may remain endemic or follow a re-emerging epidemiology dynamics which, in addition to the con-

<https://doi.org/10.1016/j.csbj.2020.07.017>

2001-0370/© 2020 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 license (<http://creativecommons.org/licenses/by/4.0/>).

tinuous public health threat of zoonotic CoVs, makes the development of new antiviral drugs crucial. Among the most promising viral targets is the Spike (S) protein, which has a critical role in viral infection and pathogenesis. It is surface exposed and mediates the entry of CoVs in the host cell, which makes it the main target of neutralizing antibodies and an obvious antiviral target for new drug development [3–7].

Two main features have been associated with the zoonotic potential of CoVs and are important determinants of infectivity, pathogenesis and host range, the receptor recognition and a multi-basic cleavage S1/S2 site for priming of the S protein conferring high-cleavability [3,8–10]. Receptor binding induces conformational changes in the S protein, exposing cleavage sites for priming proteases [11]. Binding to a receptor is mediated by a receptor-binding domain (RBD), which is located in the surface unit S1. The S1 subunit generally consists of an N-terminal domain (NTD) and a C-terminal domain, which can serve as RBD either alone or in combination [11]. For SARS-CoV-2 S protein, Angiotensin Converting Enzyme-2 (ACE2) has been identified as the primary receptor, as for SARS-CoVs who also belongs to the B lineage of betacoronavirus (beta-CoVs), but not for MERS-CoVs from the C lineage which uses human dipeptidyl peptidase 4 (DPP4) [3,4,11–14].

Priming occurs at S1/S2 interface and S2' sites, providing the CoV S protein with the structural flexibility required for the membrane fusion reaction. The cleavage at S2' generates the mature N-terminus of the fusion peptide (FP), that is inserted into the membrane [11]. In SARS-CoV-2, SARS-CoVs and MERS-CoVs the priming proteolytic cleavage process is carried out by human Transmembrane Protease Serine 2 (TMPRSS2), although MERS-CoVs require a pre-cleavage for subsequent S protein activation by TMPRSS2 carried out by furin in infected cells [4,11,15–19].

The aim of this study was to identify and map druggable consensus hot spots or regions in the three-dimensional structure of the S protein of SARS-CoV-2, that can act as antiviral targets for the development of new molecules against a broad-range of beta-CoVs. This research also contributes with a new panel for Spike structure–function studies, which can accelerate the Spike target validation and prompt in silico-chemico-biological approaches that aid in the discovery of potent antiviral drugs or monoclonal antibodies. We used a comprehensive approach, combining data from a conservation analysis of amino acid (aa) sequences from Beta-CoVs (lineages B and C), with data from a druggability study on SARS-CoV-2 crystallographic structures.

Anti-coronavirus strategies based on highly conserved and druggable targets are lacking. Structure-based design approaches have proven fundamental in the discovery and development of new drugs, including influenza antivirals; and can reduce the time and costs associated to the *de novo* drug development [20,21]. Also, predicting the druggability of the targets is a critical step in drug discovery, particularly since undruggable targets are responsible for up to 60% of failure in this process [22,23].

2. Results and discussion

2.1. Spike length and underlying variations in primary protein structure

S length varies across all four beta-CoVs. SARS-CoV-2 harbours a protein 18-aa longer than the SARS-CoVs (1273 aa); and a high inter-variability in length was observed in bat-SL-CoV S sequences (1128 to 1269 aa), concordant with the high diversity of bat-SL-CoVs found in bat populations (Table 1) [24]. MERS-CoVs harbour the longest S protein, with further 80 aa than the SARS-CoV-2 (1353 aa; Table 1). The differences in S length are entirely attrib-

uted to differences in the S1 subunit, except in the case of MERS-CoVs (Table 1). Fifteen different length variation types were found in bat-SL-CoV S1 sequences (640–681 aa range), with bat/Yunnan/RaTG13/2013 (RaTG13) and LYRa11/R.affinis/Yunnan/2011 (LYRa11) viruses harbouring the S1 subunit most similar in length to SARS-CoV-2 S1 (681 and 671 aa, respectively). RaTG13 clusters with SARS-CoV-2 in a distinct evolutionary group from other severe acute respiratory syndrome-related coronavirus (SARSr-CoVs) [5], whereas LYRa11 is a potential recombinant descended from parental lineages considered as a gap-filling virus between bat-SL-CoVs and human SARS-CoVs [25]. Overall, 5 deletions and 10 insertions of variable length distinguished bat-SL-CoV from SARS-CoV-2 S1 aa sequences. Minor differences were found between SARS-CoV and SARS-CoV-2 S1 primary structure, namely a single 4-aa insertion and 6 deletions (3–7 aa), while MERS-CoV S1 aa sequence differed from the novel virus in 19 insertions (1–13 aa) and 13 short deletions (1–4 aa). Most of the residue insertions (1/1 SARS-CoVs, 3/5 bat-SL-CoVs, and 11/19 MERS-CoVs) and most (4/6 SARS-CoVs) or half (5/10 bat-SL-CoVs, 6/13 MERS-CoVs) of the aa deletions observed take place in the S1-NTD, evidencing that this specific domain might not be a suitable target for a wide-reaching antiviral strategy against all four beta-CoVs. S1-NTD is responsible for binding sugar receptors for cell entry and differences in this domain may result in an altered specificity for a different sugar receptor [3]. In fact, NTD of middle east respiratory syndrome-related coronavirus (MERSr-CoVs) binds preferably to α 2,3-linked sialic acid, and no sugar binding has been reported for SARS-CoVsNTD [13]. Whether or not SARS-CoV-2 NTD binds to sugar remains unknown to date. At the S1/S2 boundary there is a polybasic cleavage site, between aa 682 to 685 (SARS-CoV-2 numbering), in SARS-CoV-2 (RRAR) and MERS-CoVs (RSVR) (Table 1), but not in SARS-CoVs and bat-SL-CoVs, in which the S protein remains uncleaved during assembly and exocytosis. The presence of a S1/S2 cleavage site in MERS-CoVs revealed to be important, but not essential, for S protein activation [26,27]. Whether or not this extends to SARS-CoV-2 S protein remains to be determined, leaving the path open for a potential virus-specific antiviral strategy.

The S2 subunit of SARSr-CoVs are equal in length (588 aa); whereas the MERS-CoVsS2 is 14-aa longer (602 aa; Table 1), having been identified 9 insertions (1–5 aa) and 3 short deletions (1–4 aa), compared to SARSr-CoVsaa sequence. Four of the 9 insertions lie at the central helix (CH), which may enable a slightly different conformation of the helix promoted by the binding to a different receptor [5].

2.2. Conservation studies

2.2.1. SARS-CoV-2

The S protein of the SARS-CoV-2 is highly conserved. In the S1 and S2 subunits, 99.8% and 95.6% of the residues are highly conserved, respectively (conservation score (cs) \geq 10). In S1-SD2, only position 614 was found to be variable (68.7% D614 vs 31.3% G614). There is a widespread geographic distribution of the D614G mutation since SARS-CoV-2 strains bearing this mutation were found in all continents. The majority of strains isolated in North and South America, Oceania and Asia, contain a D residue at position 614, while most of strains from Europe, Africa and Central America contain the substitution D614G. Conversely, in the S2 subunit, 10 variable residues (cs < 7) were found at positions D839, A852, V860, L861 (connecting region (CR) between FP and heptad repeat 1 (HR1) domains); S939, S940 and S943 (in HR1); V1040 (in CH); and C1254 and P1263 (in cytoplasmic domain (CP)). Specifically, the variants D839Y and D839E can be identified; the variant A852V occurs in strains from Netherlands; V860Q and L861K co-appear in the same strain from Beijing; S939F occurs exclusively

Table 1 Distinctive features (length variation and cleavage site) and number of sequences enclosed in the datasets of the Spike S1 and S2 subunits of each betacoronavirus analysed.

Genus	Subgenus	Species	Lineage	Virus	SPIKE PROTEIN (S gene)														
					Whole-protein					S1 subunit					S2 subunit				
					Position in the genome (nts) ^{a,b}	Length nts	Length aa	Position in the genome (nts) ^a	Length nts	Length aa	Number of sequences	Position in the genome (nts) ^b	Length nts	Length aa	Number of sequences	Position in the genome (nts) ^b	Length nts	Length aa	Number of sequences
Beta-CoV	Sarbecovirus	SARSr-Cov	B	SARS-CoV-2	21314–25135	3822	1273	21314–23368	2055	685	674	23369–25135	1767	588	682	682 ^c RRAR ⁶⁸⁵			
		SARSr-Cov		SARS-CoV	21228–24995	3768	1255	21228–23228	2001	667	114	23229–24995	1767	588	116	-			
		SARSr-Cov		bat-SL-CoV	21,264 - [24950–25073]	3687–3810	1228–1269	21,264 - [23183–23306]	1920–2043	640–681 ^c	50	[23184–23307] - [24950–25073]	1767	588	50	-			
Merbecovirus	MERSr-Cov	C		MERS-CoV	21178–25239	4062	1353	21178–23430	2253	751	248	23431–25239	1809	602	248	682 ^c RSVR ⁶⁸⁵			
					TOTAL	1086	TOTAL	1096											

^a counting from the start codon of the first genome-encoded protein (nsp1); ^b includes the STOP codon; ^c 15 different lengths: 649 (15/50); 648 (9/50); 653 (7/50); 663 (5/50); 640, 668, and 667 (2/50); 681, 662, 657, 658, 647, 654, 646, and 671 (1/50); ^d SARS-CoV-2 numbering

in strains from Europe; S940F and C1254 occur in Australia isolates; S943P and S943T variants are found in Belgium isolates; V140F appears in strains from China; and P1263L occurs in strains from England and Henan. The [supplemental Figure S-1a](#) displays the SARS-CoV-2 conservation-score distribution for the S S1 and S2 subunits.

The S-RBD, described in the literature as a promising antiviral target [28–30], contains conserved and highly conserved residues, presenting only minor aa variations with equivalent aa properties: five strains isolated in France present V367F and four strains isolated in the USA bear the aa substitution V483A.

2.2.2. SARSr- and MERSr-CoVs

The S protein is mostly conserved among SARSr-CoVs (SARS-CoV-2, SARS-CoVs and bat-SL-CoVs). In regard to human SARSr-CoVs (hSARSr-CoVs), 64.7% of the residues within the S1 subunit are highly conserved (cs ≥ 10); 11.2% are conserved (cs of 7–9) and only 24.1% are considered variable (cs ≤ 6.9). Most variable sites/residues are found in S1-NTD and at the C-terminal end of SD2. The S2 subunit is mostly highly conserved: 90% of the residues are highly conserved, 2.5% are conserved and only 7.2% are variable (mostly located in the S2-NTD and in the CR. A similar conservation pattern was found in bat-SL-CoVs.

In regard to SARSr- and MERSr-CoVs (SARS-CoV-2, SARS-CoVs, bat-SL-CoVs and MERS-CoVs), in the S1 subunit 20.2% of the residues are highly conserved, 36.7% are conserved, and 43.1% are variable. In the S2 subunit, 42.2% of the residues are highly conserved, 23.8% are conserved and 34% are variable.

Variable sites/residues are found in SARS-CoV-2, SARSr-CoVs, and between SARSr- and MERSr-CoVs. It possibly suggests strong selective pressure; and, in this context, these residues or regions do not constitute, *a priori*, suitable regions for the design of a promising antiviral strategy targeting the S protein. The Figure S-1 displays the S conservation-score distribution for the hSARSr-CoVs (Figure S-1b) and for the SARSr- and MERSr-CoVs (Figure S-1c).

The conservation scores for the hSARSr-CoVs and the SARSr- and MERSr-CoVs were mapped onto the S monomeric structure (SARS-CoV-2 structure, PDB entry 6VYB) as illustrated in Fig. 1.

Additionally, we have identified conserved clusters (CC) (≥10 continuous residues) on the S S1 and S2 subunits of hSARSr-CoVs and between SARSr- and MERSr-CoVs (Table 2; Figure S-1). High level of conservation was also found between SARS-CoVs, SARS-CoV-2 and bat-SL-CoVs, with CC overlapping the ones described for hSARSr-CoVs (Table 2).

The hSARSr-CoV S1 subunit contains 15 CC, named according to the dominant domain of the clustered residues (listed in Table 2). The longest S1 CCs consist of RBD.1, RBD.4 and SD1.1. The CCs RBD.3 and RBD.5 exhibit a conservation score of > 10.5 for the hSARSr-CoVs and the NTD.4, RBD.3, SD1.1 and SD2.2 are also conserved among all beta-CoVs included in the study.

The hSARSr-CoVs S2 subunit comprises 14 CC (Table 2). The CC S2-NTD.2, FP.1, CR.1, CR-HR1.1, HR1-CH.1, CH.1, heptad repeat 2-cytoplasmic domain.1 (HR2-CP.1) and CP.1 exhibit a conservation score of >10.5 for the hSARSr-CoVs. The longest S2 CCs consist of HR1-CH.1 and connector domain (CD)-HR2.1. Some of these CC are concurrently identified as shorter variant (sv) forms, when comparing hSARSr-CoVs and between SARSr- and MERSr-CoVs. Specifically, a shorter variant of CR.1 (sv-CR.1) and CR-HR1.1 (sv-CR-HR1.1), three shorter variants of HR1-CH.1 (sv1-HR1-CH.1, sv2-HR1-CH.1 and sv3-HR1-CH.1), and two shorter variants of CD-HR2.1 (sv1-CD-HR2.1 and sv2-CD-HR2.1) are found among all beta-CoVs included in the study.

Taken all together, high level of conservation was found in several domains of S1 and S2 subunits of hSARSr-CoV S, with a similar number of CC found in each subunit. The majority of the S1 CC are

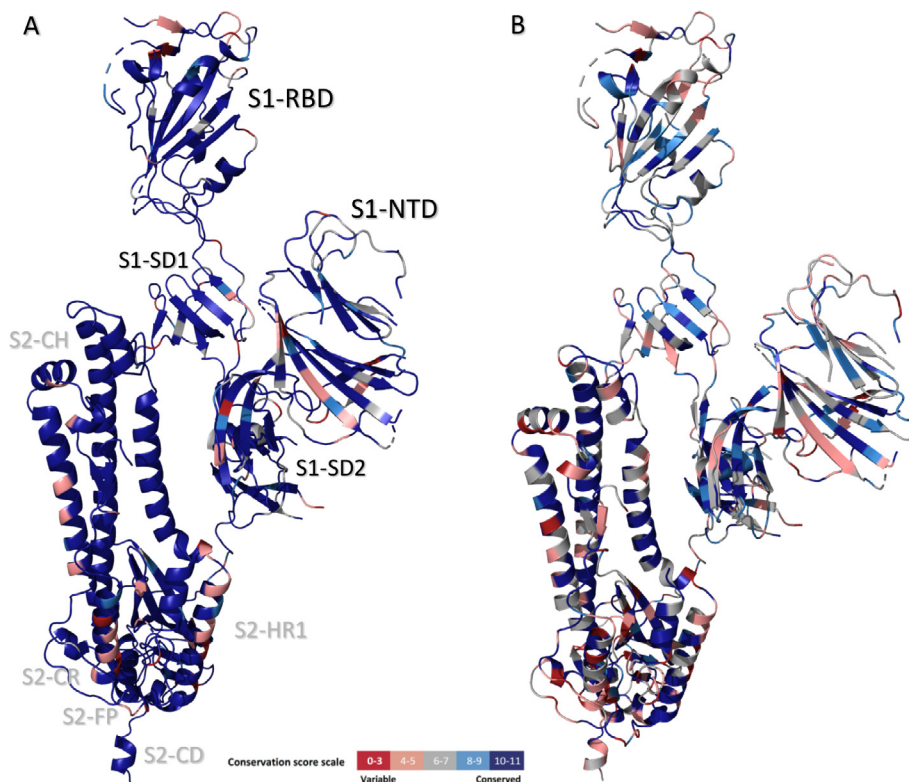


Fig. 1. Three-dimensional sequence conservation of the human SARSr-CoVs (A) and SARSr- and MERSr-CoVs (B) Spike protomers (PDB ID: 6VYB). The cs are mapped onto the S protein structures. The figure is coloured based on cs: the highest conservation positions are highlighted in shades of blue and the red regions indicate low conservation according to the cs scale. The S1 and S2 domains resolved are indicated in black and grey, respectively, in the hSARSr-CoV protomer. NTD: N-terminal domain; RBD: receptor-binding domain; SD: subdomain; FP: fusion peptide; CR: connecting region, HR: heptad repeat; CH: central helix. The figures were produced with PyMOL Molecular Graphics System [101]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

located in the RBD, while the S2 subunit comprises longest CC and mostly located in the protein regions that encompasses the CR, HR1, CH and HR2 domains. The average degree of protein sequence identity for some of these S2 CC is approximately 100%, which corresponds to almost absolute conserved sites ($cs > 10.5$).

In regard to SARSr- and MERSr-CoVs, 4 CC were found in the S1 subunit and 7 CC were found in the S2 subunit. The largest pocket within S1 lies in the RBD and corresponds to the RBD.3 described for hSARSr-CoVs. The majority of the S2 CC are located in the protein regions that comprises the domains CR, HR1 and CH.

Overall, this suggests that the S1 subunit exhibits a slightly highest conservation (99.8%) when compared to S2 (95.6%) in SARS-CoV-2. However, for hSARSr-CoVs and when comparing SARSr- and MERSr-CoVs, S2 subunit presents a higher level of conservation.

A good antiviral target will have a structure predominantly conserved, with advantageous structural/druggable features, and concomitantly displays an important function regarding the virus replication cycle [31,32]. We identified the S1-RBD and the regions that include the CR, HR1 and CH in the S2 subunit, as the most promising domains and further explored them for druggability.

2.3. Druggability studies

Only pockets comprised of a minimum of 10 aa and defined with a druggability score ≥ 0.4 were considered (as described in the Materials and Methods section), based on the premise that a minimal pocket size is required in order to indulge a proper interaction between the target and potential ligands. Pockets with an impaired ability to bind a ligand are named decoy pockets and frequently include small cavities comprised by less than 10–14 resi-

dues [33,34]. Druggable pockets are required to bear other advantageous structural features which can be estimated/characterized using pocket descriptors that include the volume, depth, enclosure, solvent-accessible surface area, buriedness of residues in the interface, etc [35].

In the S pre-fusion state, three monomers intertwine to form a trimer complex which represents the predominant conformation of the S protein in vivo [6]. The binding of the S protein to the receptor promotes a conformational switch of the S1-RBD from down to up conformation (determining S closed and open states). This results in a dissociation of the S1-ACE2 complex and S1 monomers from the pre-fusion protein and prompts the conformational transition to the post-fusion state. The specific mechanism, and the range of S conformational changes and states, remain uncharacterized. [36–38]. Several studies have shown that the ectodomain and the RBD expression in eukaryotic systems are stable and appear in the monomer conformation in high yield; and that soluble ACE2 binds to both monomer and trimer conformations [37,39–42]. In this context, to fully characterize the S protein, it is important to explore both the monomer and trimer in structural studies. A comprehensive analysis of the full-length S monomer (in both open and closed conformations) is discussed in a section of the Supplemental Data.

2.3.1. Spike-RBD

The S-RBD has been repeatedly emphasized by several authors as bearing a higher antiviral target potential [28–30], and, for that reason, we describe it in more detail. Sixteen druggable pockets were identified in the three S-RBD structures (from a total of 29) for the DoGSiteScorer (DGSS) server. Most pockets are shared by all the RBD structures, since no major conformational differences

Table 2

Summary of the conserved regions identified in the S1 and S2 subunits of the Spike protein. Only conserved clusters comprised by a minimum of 10 aa and with a conserved score ≥ 7 were considered for either the hSARSr-CoVs, the SARSr-CoVs (including the bat-SL-CoVs) and the SARSr- and MERSr-CoVs. Regions common to all three groups are highlighted in bold and blue.

human SARSr-CoV						SARSr-CoV						SARSr- and MERSr-CoV					
S1 subunit			S2 subunit			S1 subunit			S2 subunit			S1 subunit			S2 subunit		
Region	Position	Lenght (aa)	Region	Position	Lenght (aa)	Region (analogue to the hSARSr-CoVs)	Position	Lenght (aa)	Region	Position	Lenght (aa)	Region (analogue to the hSARSr-CoVs)	Position	Lenght (aa)	Region	Position	Lenght (aa)
NTD.1	T51-T63	13	NTD.1	S691-E702	12	NTD.1	T51-T63	13	S2-NTD.1	S691-E702	12	NTD.4	V308-V320	13	sv-CR.1	I818-K835	18
NTD.2	T95-G107	13	NTD.2	M740-C749	10	NTD.2	T95-G107	13	S2-NTD.2	M740-C749	10	RBD.3	S373-L390	17	sv-CR-HR1.1	Y904-Q913	10
NTD.3	T274-C291	18	NTD.3	N751-L767	17		A262-P272	11	S2-NTD.3	N751-L767	17	SD1.1	C538-T553	16	sv1-HR1-CH.1	L966-S975	11
NTD.4	V308-V320	13	FP.1	F797-P809	13	NTD.3	T274-C291	18	FP.1	F797-P809	13	SD2.4	S659-A672	14	sv2-HR1-CH.1	E990-L1001	12
RBD.1	I326-F347	22	CR.1	K814-F833	20	NTD.4	V308-V320	13	CR.1	K814-F833	20	-	-	-	sv3-HR1-CH.1	T1006-A1015	10
RBD.2	R355-A372	18	CR.2	K835-I844	10	RBD.1	I326-F347	22	CR.2	K835-I844	10	-	-	-	sv1-CD-HR2.1	S1147-N1158	12
RBD.3	F374-S383	10	CR.3	A846-A871	26	RBD.2	R355-A372	18	CR.3	A846-A871	26	-	-	-	sv2-CD-HR2.1	K1211-M1229	19
RBD.4	N394-G416	23	CR-HR1.1	G885-K921	37	RBD.3	F374-G416	43	CR-HR1.1	G885-N928	43						
RBD.5	I418-F429	12	HR1-CH.1	A944-Q1054	111	RBD.4	I418-F429	12	HR1-CH.1	A944-Q1054	111						
RBD.6	G502-L518	17	CH.1	A1056-P1069	14	RBD.5	I418-F429	12	CH.1	A1056-P1069	14						
SD1.1	S530-T553	24	CH-CD.1	Q1071-V1096	26	SD1.1	L533-T553	21	CH-CD.1	Q1071-V1096	26						
SD2.1	P589-S605	17	CD-HR2.1	P1112-I1232	121	SD2.1	P589-S605	17	CD-HR2.1	P1112-I1232	121						
SD2.2	A623-G639	17	HR2-CP.1	L1234-G1246	13	SD2.2	A623-G639	17	HR2-CP.1	L1234-G1246	13						
SD2.3	N641-N657	17	CP.1	C1248-T1273	26	SD2.3	N641-N657	17	CP.1	C1248-T1273	26						
SD2.4	S659-T676	18	-	-	-	SD2.4	S659-T676	18	-	-	-						

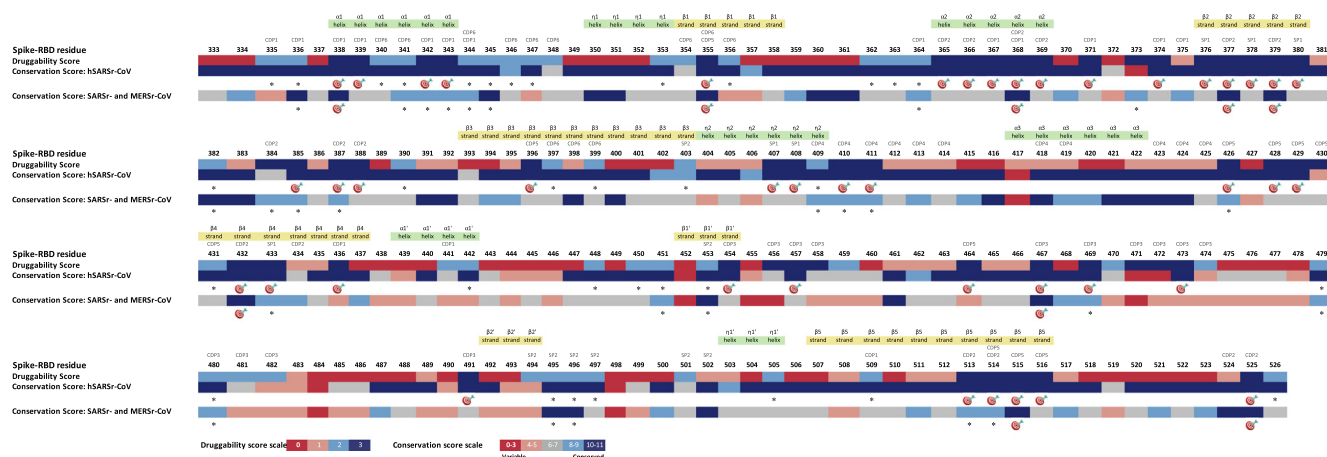


Fig. 2. Overall alignment of Spike-RBD druggability along with the conservation scores for each residue position. The druggability prediction was based on the descriptors algorithm of each pocket bioinformatics tool: SF, DGSS and PDS. The potential conserved druggable sites/residues are marked with an asterisk and the T-RHS are marked with a target. The conserved druggable pockets allocated to each site/residue along with the secondary structure elements are indicated at the top of the picture.

exist among them. The [supplement Table S-4](#) displays the 16 pockets identified for each structure, along with the pocket descriptors: size, volume and druggability score.

A linear schematic representation of the S-RBD containing the druggability information for each aa position is shown in [Fig. 2](#). Differences in the distribution of druggable aa regions/residues within the RBD, may help to distinguish which ones have a potential role in RBD function or structure. The conserved druggable regions and the top-ranked hot spots (T-RHS) identified through comparative analysis of the S-RBD druggability and conservation scores in both hSARSr-CoVs and SARSr- and MERSr-CoVs are depicted in [Fig. 2](#).

Some of these conserved druggable regions are located close together and can form larger pockets, named as consensus druggable pockets (CDP). Six CDP (common to all RBD structures) were characterized based on the comparative druggability study. The secondary structure elements were predicted based on the ESPript, as previously performed by Wang *et al.* [43,44]. Three out of these 6 CDP have more than 14 aa and, interestingly, one of the CDPs is not found in the most complex structures (monomer or trimer). The [supplemental Table S-2](#) displays the CDPs identified for the S-RBD (along with the aa number and composition).

The CDP1 (helices $\alpha 1$ - $\alpha 2$, $\alpha 1'$ and strands $\beta 4$ - $\beta 5$) is comprised of 16 residues. The CDP2 (helix $\alpha 2$ and strands $\beta 2$, $\beta 4$ and $\beta 5$) comprises 15 residues. The CDP3 (strands $\beta 1'$ and $\beta 2'$), only found in S-RBD structures, contains 14 residues. The CDP4 (helix $\eta 2$ and $\alpha 3$) is comprised by 11 residues; the same number for the CDP15 (strands $\beta 1$, $\beta 3$ - $\beta 5$). The CDP6 (helix $\alpha 1$; strands $\beta 1$ and $\beta 3$) con-

tains 12 residues. Two small pockets (SP1 and SP2) comprised by 8 residues each were found in all RBD structures. The spatial arrangement of the highest-ranked CDPs for the S-RBD based on the DGSS algorithm is shown in [Fig. 3](#).

The 43T-RHS for hSARSr-CoVs and 9 for SARSr- and MERSr-CoVs (out of 75 and 35 conserved druggable residues - CDR, respectively) are listed in [Table S-3](#) and depicted in [Fig. 2](#); and demonstrate that the S-RBD bears a high potential of druggability. CDR and T-RHS have been identified for hSARSr-CoVs and for SARSr- and MERSr-CoVs, establishing S-RBD as a promising anti-CoV target for the design and development of antiviral and vaccine strategies.

2.3.2. Spike trimer conformation

The homotrimeric S protein is suggested to adopt two dynamic distinct states (referred to as closed and open conformation). In the S trimer analysis, both closed and open state conformations along with an asymmetrical homotrimer, in which a single RBD domain is featured in the up conformation (semi-open state), were included.

A total of 170 potential druggable pockets (out of 304) were identified in both open (61 pockets), semi-open (67 pockets) and closed (42 pockets) conformations, for the DGSS server. Several pockets are shared between the three structures, independently of the RBD conformation. Moreover, it is possible to identify pockets common to the CDPs previously predicted for the S monomer structures, however some are exclusively found when the S protein assembles into a homotrimeric complex. The [supplemental](#)

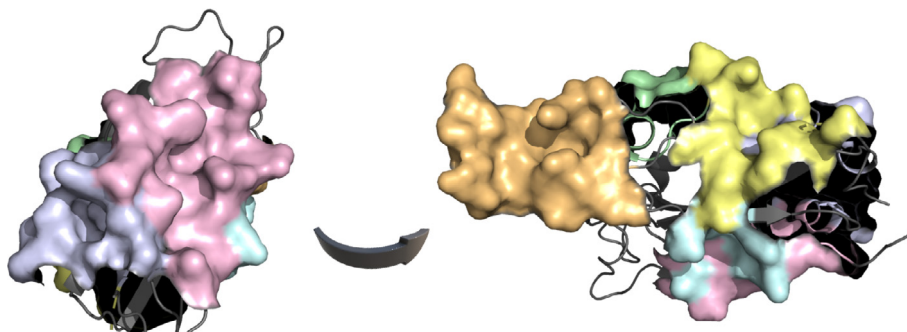


Fig. 3. Mapping results of the highest-ranked consensus druggable pockets onto the Spike-RBD crystallographic structure of SARS-CoV-2 (PDB: 6VW1). The top-ranked pockets: CDP1 to CDP5 are highlighted in shades of pink, pale blue, orange, green, yellow and cyan, respectively. The number and aa composition of each CDP are shown in [Supplemental Table S-2](#). The figures were produced with PyMOL Molecular Graphics System [101]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table S-5 displays the pockets identified for each S trimer conformation, along with the pocket descriptors: size, volume and druggability score.

Similarly, several pockets identified in the trimer are also shared by all structures, independently of the conformational RBD switch within the structure. The CDPs previously identified for the S monomer (CDP.M) maintain advantageous structural features regarding the druggability prediction in the trimeric structures (CDP.T). None was found to be surface-inaccessible and the residues were not found to be buried in the interface when the monomer subunits combine to form a trimeric complex, thus, all CDPs identified for the S monomer group are represented in the trimer.

28 CDPs (22 CDP with ≥ 14 aa) were found in S trimer structures, 12 in S1 (CDP1T-S1 to CDP12T-S1), 11 in S2 (CDP1T-S2 to CDP11T-S2) and 5 incorporate both S1 and S2 (CDP1T-S1S2 to CDP5T-S1S2). The CDPs identified for the S trimer can be composed of one, two or the three monomer subunits of the trimeric complex. Some of these CDPs have not been identified in the monomer analysis, suggesting that the trimeric structure contributes to expand and improve the druggability potential of the protein. Thus, a higher number of CDPs (and also T-RHS and CDR, as described below) can be found in the trimer conformation in comparison to S monomer.

Four S1 CDPs (CDP3T-S1, CDP5T-S1, CDP9T-S1 and CDP11T-S1), located mainly in the RBD, can only be found in the trimeric structure. In the same vein, 7 pockets identified within the S2 subunit (CDP2T-S2, CDP4T-S2, CDP5T-S2, CDP6T-S2, CDP8T-S2, CDP9T-S2 and CDP11T-S2) and all the 5 CDPs which incorporate both S1 and S2 subunits (CDP1T-S1S2 to CDP5T-S1S2) are only represented in the trimer. The supplemental Table S-2 displays the CDPs identified for the S trimer conformation (along with the aa number and composition; and the corresponding location/domain).

The 22 CDPs shared by all trimeric structures and exhibiting ≥ 14 aa have been analyzed according to the sequence location within the protein (from the N- to the C-terminal). In regard to the S1 subunit, the first two CDPs (CDP1T-S1 and CDP2T-S1) are located in the S1-NTD and are very similar in structure and aa composition to the CDP1M-S1 and CDP2M-S1, respectively, described

for the monomer conformation. The CDP3T-S1 (in the NTD and RBD) presents as a larger CDP that resembles a combination of both CDP3M-S1 and CDP11M-S1. The CDP4T-S1 (in the NTD) is analogous to the CDP5M-S1. The CDP5T-S1 (in the NTD and RBD) is only found when two monomer subunits bind together. The CDP6T-S1 (in the NTD) and the CDP7T-S1 (in the RBD) are analogous to the CDP6M-S1 and CDP8M-S1, respectively. The CDP9T-S1 and CDP11T-S1 (in the RBD) are only described for the trimeric structure. The CDP12T-S1 (in the NTD) is analogous to the CDP14M-S1.

In regard to the S2 subunit, the CDP1T-S2 is analogous to the CDP1M-S2 described for the monomer structure. A larger pocket composed of 163 residues, CDP2T-S2, is only found in the trimeric structure when three monomers bind together; and is similar in composition to the CDP1T-S2 and the analogous pockets CDP3M-S2 and CDP9M-S2. The CDP3T-S2 is analogous to the CDP4M-S2. The CDP8M-S2 described in the monomer analysis can be found within the trimeric structure in a more complex (recruiting \geq two monomer subunits) larger pocket: (1) either together with the CDP6M-S2 to form the CDP4T-S2; or (2) together with both CDP6M-S2 and CDP7M-S2 to form the CDP5T-S2. The CDP6T-S2 (in the S2-NTD and CH) is only found in the trimeric structure when two protomers bind together. The CDP7T-S2 is analogous to the CDP2M-S2. The CDP8T-S2 (in the S2-NTD, FP, HR1, CH and CD) is only identified in the trimeric structure. The CDP9T-S2 (mainly located in the HR1, CH and CD) is described only in the trimeric structure. The CDP10T-S2 is analogous to the CDP7M-S2.

In addition to the CDPs described for S1 and S2, the trimeric structure bears CDPs that include both subunits. The CDP1T-S1S2 (of 40 residues) is in part similar to the CDP3M-S2 described for the monomer structure (located in the S2-NTD, HR1 and CH) but is also composed of additional residues pertaining to the S1-NTD. The CDP2T-S1S2 (of 71 residues) is formed by a pocket pertaining to two protomers binding together, one analogous to the previously described CDP1T-S1S2 and the other composed of residues that form both the CDP12T-S1 and the CDP13M-S1. The CDP3T-S1S2 (of 63 residues) is in part analogous to the CDP7M-S1 (in the RBD); while it also comprises additional residues belonging to the SD1 and S2-HR1. The CDP4T-S1S2 is located in the S1-RBD



Fig. 4. Mapping results of the highest-ranked consensus druggable pockets onto the Spike trimer crystallographic structure of SARS-CoV-2 (PDB: 6VYB). The top-ranked pockets: CDP5T-S1, CDP6T-S1, CDP9T-S1, CDP2T-S2, CDP4T-S2, CDP8T-S2, CDP2T-S1S2, CDP3T-S1S2, CDP5T-S1S2 are highlighted in shades of yellow, turquoise, pale blue, cyan, orange, pink, grey, green and raspberry, respectively. The number and aa composition of each CDP are shown in Supplemental Table S-2. The figures were produced with PyMOL Molecular Graphics System [101]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

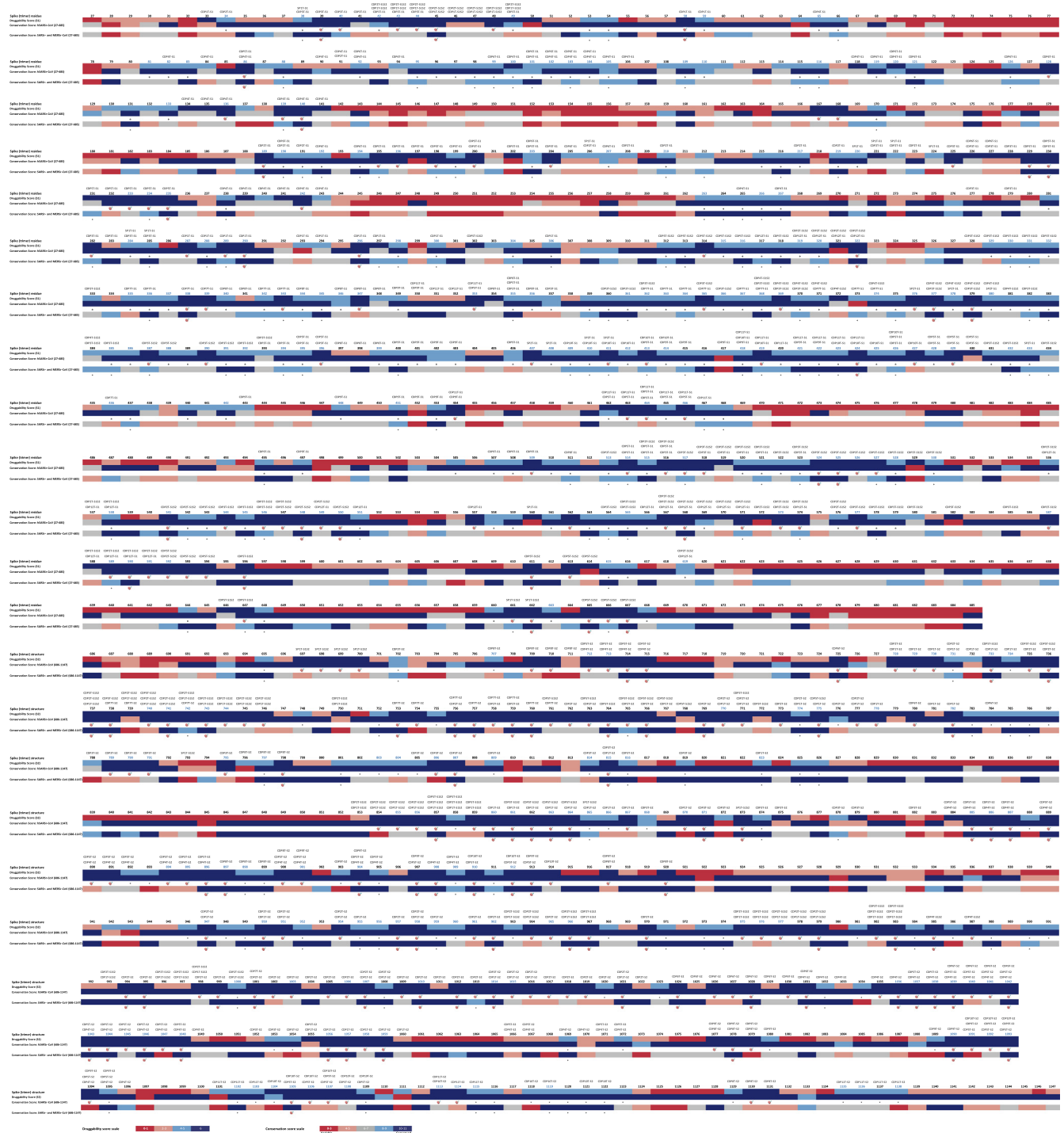


Fig. 5. Overall alignment of Spike S1 and S2 druggability along with the conservation scores for each residue position. The druggability prediction was based on the descriptors algorithm of each pocket bioinformatics tool: SF and DGSS; and for all the predicted S state conformation (open, semi-open and open). The potential conserved druggable sites/residues are marked with an asterisk and the T-RHS are marked with a target. The conserved druggable residues shared by the S monomer and trimer conformations for the hSARS-CoV are coloured in blue; and the conserved druggable pockets allocated to each site/residue are indicated at the top of the picture. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and S2-CH. The CDP5T-S1S2 (of 42 residues) is also formed by a pocket pertaining to two protomers binding together, one (19 aa) located in the S1-NTD and S2 and the other (23 aa) located in the S2-NTD.

Although the analysis of the differences between S state conformations is out of the scope of this article, two main differences are worth to mention. First, a pocket of 20 residues located in the S1-NTD and RBD (aa 117, 128–130, 167–170, 229–231, 357, 359–360, 393–394, 520–521, 523), with no expression in the closed S confor-

mation, is found in the open S conformation structures. Second, a large pocket (in the RBD) comprised of the combination of CDP3T-S1, CDP8T-S1, CDP9T-S1, CDP10T-S1, CDP11T-S1, CDP3T-S1S2, CDP4T-S1S2 and SP2T-S1 is only found in the open conformation structure. Taken together, we suggest that, for the trimeric complex, the open conformation bears additional potential binding regions within the RBD for distinct modes or types of S-RBD interactions. The spatial arrangement of the highest-ranked CDPs for the trimer conformation is shown in Fig. 4.

The last stage of the study consisted of a global comparative analysis of the S trimer druggability together with the conservation scores for the hSARSr-CoVs and the SARSr- and MERSr-CoVs (Fig. 5).

The 274T-RHS for drug targeting (out of 553 CDR) identified in the hSARSr-CoV S trimer comprise 99 hot spots in S1 (out of 288 CDR); and 175 hot spots (out of 265 CDR) in S2. The [supplemental Table S-3](#) displays the CDR and T-RHS identified for the S trimer conformation. Comparing the two subunits, no major differences were found in the absolute number of CDR. S1 displays a slightly higher number of CDR compared to S2 (S1: 288 vs S2: 265). However, most of the T-RHS are located within S2, with a prevalence of 29.8% (175/588), whereas S1 shows a prevalence of 14.4% (99/685). It has been found that about 42% of the CDR in S1 are located in the RBD. The RBD is the most druggable domain within the S1 subunit (120 CDR and 33T-RHS out of 195 overall residues), followed by the SD1 (with 38 CDR and 19T-RHS of a total of 94 residues). The SD2 is the least relevant domain within the S1 subunit in regard to the druggability potential, as previously predicted in the monomer analysis. For the S2 subunit, it is worth to mention that 12 CDR were found within the FP (total length of 19 aa); and CH was identified as the most conserved druggable region (65 CDR and 50T-RHS in a 99-aa domain length) although the FP, S2-NTD, HR1 and CR are also promising druggable regions (in decreasing ranking order).

The 110T-RHS for drug targeting (out of 164 CDR) found in the SARSr- and MERSr-CoVs comprise 28 hot spots (out of 160 CDR) in the S1 subunit and 82 hot spots (out of 132 CDR) in the S2 subunit (Table S-3). In the same vein to that observed for the hSARSr-CoVs, S1 displays a higher number of CDR compared to S2 (S1: 160 CDR vs S2: 132 CDR), while most of the T-RHS are located within S2 (S1: 28 vs S2: 82). The majority of the S1 CDR are located in the NTD, RBD and SD1. The SD1 displays 20 CDR and 5T-RHS out of 64 residues that comprise the domain. Although the RBD was found to be the most druggable domain in hSARSr-CoVs, it ranks second in the SARSr- and MERSr CoV group, with some regions within the RBD showing lower conservation among all four beta-CoVs. The S1-SD1 is the most conserved druggable domain for SARSr- and MERSr-CoVs, followed by the RBD. For the S2 subunit, a different pattern is evident when compared to hSARSr-CoVs. Most of the CDR or T-RHS are found in the CH, followed by the HR1 and CR. The FP displays a lower conservation score among all CoV species and does not represent a conserved druggable domain in the SARSr- and MERSr-CoV group. The CH is the most conserved druggable domain in the S2 subunit, for this group.

The comparative analysis of the S monomer and trimer conformations highlighted that 171 S1 CDR and 119 S2 CDR are shared by these conformations, in regard to the hSARSr-CoVs. Moreover, 95 S1 CDR and 59 S2 CDR are present in both monomer and trimer conformations in SARSr- and MERSr-CoVs (Table S-6). The CDR shared by the monomer and trimer conformations are highlighted in blue in Fig. 5.

2.4. Contribution of the findings to previous research

To the best of our knowledge, there are no published studies to date addressing and comparing the conservation and druggability of the CoV S protein, for such a wide range of sequences ($n = 1086$ S1; $n = 1096$ S2) from four Beta-CoVs (SARS-CoV-2, SARS-CoVs, MERS-CoVs and Bat-SL-CoVs) and the crystallographic structures of all available SARS-CoV-2 S proteins (S-RBD, monomer and trimer structures in either closed, semi-open and open state conformations, when applied). In the majority of recent studies, a comparative analysis of the S protein has been performed with one reference strain for each CoV type and thereby taking into account only the most prevalent residue harbored at a given posi-

tion; It does not represent diversity and it overestimates the protein conservation score. To overcome this, we performed the conservation analysis using the total number of protein sequences treated for each CoV type, so that the conservation estimation takes into account the variations in the aa composition within each CoV.

The present study has revealed the most propitious S domains to target in regard to the conservation and druggability analysis of both S monomer and trimer conformations. The S1-RBD represents a promising anti-COV target and is the most conserved druggable domain in the monomer analysis for hSARSr-CoVs and for SARSr- and MERSr-CoVs; and in the trimer analysis for hSARSr-CoVs. For the SARSr- and MERSr-CoV trimer, the RBD ranks second after the SD1 domain; which is concordant with different CoVs using distinct host receptors for entry [45]. The SD1 stands as the most conserved druggable domain among all four Beta-CoVs analyzed. In this context, both RBD and SD1 domains should be further addressed in future studies that target the S1 subunit.

In regard to the S2 subunit, the FP was found to bear a high potential of druggability exclusively in hSARSr-CoVs (its conservation degree decreases when considering other beta-CoV species). On the other hand, the CH demonstrates a higher conservation-druggability potential among hSARSr-CoVs and in SARSr- and MERSr-CoVs and is the most conserved druggable domain within the S2 subunit. Other S2 domains, such as the CR and HR1, are alternative potential antiviral targets and can also be considered in anti-CoV strategies.

We have demonstrated that regardless of the S protein conformation states, highly conservation regions among either the hSARSr-CoVs and the SARSr- and MERSr-CoVs can overlap with potential binding sites/residues, rendering the Spike protein a suitable antiviral target.

Our computational analysis has revealed specific T-RHS and CDR and the corresponding conserved druggable pockets within each domain of the S protein, in hSARSr-CoVs and SARSr- and MERSr-CoVs. The majority of the T-RHS identified in our study, in regard to the S-RBD analysis, lie at the core structure of this domain rather than at the receptor-binding motif (35/43 in hSARSr-CoVs; 8/9 in SARSr- and MERSr-CoVs). Recently, Wang et al. have shown that SARS-CoV-2 and/or SARS-CoV neutralizing human monoclonal antibodies 47D11 and CR3022 preferentially target the RBD core domain [46], which supports our findings. Additionally, Pinto et al. identified a human monoclonal antibody S309 with neutralizing activity against the SARS-CoV-2, which mainly targets the RBD epitope N343 [47], which is also a potential hot spot for drug targeting identified in our study. Of the S protein targeted B- and T-cell epitopes that are promising candidates for vaccine design against SARS-CoV-2 [48,49], 40B-cell and 44T-cell were identified as T-RHS or CDR residues in our analysis.

The CDR shared by the S monomer and trimer structures: C336, C361, C379, C391, C432, C525 (RBD), Q920 N955 (HR1) have been previously described in the literature for their role in protein structure [12,50]. The cysteine residues found in the RBD can form pairs of disulfide bonds (C336–C361, C379–C432 and C391–C525) that help to stabilize the β sheet structure of this domain [12]. Additionally, Lan et al. revealed a total of 17 residues from the SARS-CoV-2 RBD that are in close contact to the ACE2 receptor, but, of these, only Y453 and Y505 are highlighted by our study as trimer CDR for drug targeting [12].

Two S protein motifs within the S2 subunit have been studied by other authors. The motif KRFSFIEDLLFNKV (aa 814–826), located within and flanking the S2' cleavage site (underlined R), is suggested to be required for virus activation and cell entry [51]. Four T-RHS or CDR identified here are among the 13 residues that comprise this motif, namely: Q814, R815 (S2' cleavage site), S816 and F823. Also, the 3C-like proteinase cleavage site/region

TGRLQ[^]SLQTY (aa 998–1007), located at the S2-CH domain, is currently being studied. The majority of the residues that comprise this motif (9 out of 10) constitute T-RHS or CDR identified in our study, in particular T998, G999, R1000, L1001, Q1002, S1003, Q1005, T1006 and Y1007 [52].

The trimer-binding interface between individual S protomers and the interaction sites of trimerization are not fully characterized for the SARS-CoV-2. Recent studies have suggested several residues that contribute to the formation or to stabilize the S trimeric structure; and most of them are located at the S2 subunit (mainly in the S2-NTD, FP, CR, HR1, CH and CD) [53–58]. Thirteen of these residues (E702, Y707, N709, N710, Y789, K790, K795, F797, G798, T859, G891, Q895, F898) are described by multiple authors and represent CDR or T-RHS for drug targeting highlighted in our study [53–58]. The main CDPs that incorporate the residues in the trimer-binding interface consist of: CDP3T-S2, CDP4T-S2, CDP5T-S2, CDP8T-S2, CDP9T-S2. Additionally, Peters et al. identified 3 residues (A520, P521 and A522 – highlighted as CDR in our study) that play a role in stabilizing the RBD through interactions with the NTD of the adjacent protomers [57]. The key residues that contribute to S trimerization can be potentially modulated by therapeutic agents in order to disrupt the quaternary structure assembly of the protein.

In silico studies suggest that the SARS-CoV-2 S S1 might potentially bind to the human MERS-CoV receptor dipeptidyl peptidase-4 (DPP4, also known as CD26) [52]. Vankadari et al. predicted 14 residues within the RBD that may lie in the S1:CD26 interaction interface, but only 5 of these residues (R408, Q409, D467, S469, P491) are considered CDR based on the criteria defined in our study [52]. Furthermore, the toll-like receptor 4 (TLR4), which is involved in the recognition of molecular patterns and mediate inflammatory responses, may also interact with the S protein via 10 residues located in the S1 subunit. Three of these residues were identified in this study as T-RHS residues, namely Y204, V289 (S1-NTD) and F562 (SD1). Both the S:CD26 and S:TLR4 interactions may constitute potential alternative broad antiviral targets [59].

It is established that SARS-CoV-2 S protein enters into the host cell through the ACE2 receptor; but it also uses sialic acids linked to gangliosides at the host plasma membrane, which may improve the virus attachment to lipid rafts and facilitate the contact with the ACE2 receptor [60]. Fantini, et al. have identified a ganglioside-binding domain at the S1-NTD of S protein (aa 111–162) [60]. The residues S116, I119, V120, N121, V126, I128, F133, C136, P139, F140 comprise the ganglioside-binding domain and were identified in our study as potential T-RHS or CDR for drug targeting. Moreover, chloroquine, and its close structural analogues, bind sialic acids and gangliosides with high affinity and have shown to block the S:ganglioside interaction [60]. The authors also suggested that the azithromycin might interact with this ganglioside-binding domain within the S protein [61].

Drug repurposing or the chemical optimization of existing drugs represent an effective drug discovery approach which has the potential to reduce the time and costs associated to the *de novo* drug discovery and development and the subsequent clinical trials process [62]. *In silico* and *in vitro* studies have recently demonstrated that the arbidol (anti-influenza inhibitor) and the nelfinavir mesylate (anti-HIV inhibitor) can potentially inhibit the SARS-CoV-2 replication [63,64]. The majority of the key residues involved in the arbidol:S interaction (7 out of 9 residues) have been identified here as potential hot spots, in particular: E780, K947, E1017, R1019, S1021, L1024, T1027 (S2-NTD, HR1, and CH domains). In regard to nelfinavir, Musarrat et al., have shown that the following CDR may lie in the interface of S:nelfinavir interaction: Q954, Q957, A956, L1012, I1013 (FP and HR1 domains) [64]. Other docking assays have suggested that the CDR: R319 (S1-NTD), C391, L517 (RBD), C538, F543, N544, Q564, P589 and S591 identified in

this study represent active site residues that potentially interact with several medicinal compounds such as arzanol, genistein, resveratrol, rosmanol or thymohydroquinone [65].

The HR1 has also been identified in our study as a promising conserved druggable region. Based on the finding that HR1 and HR2 regions are able to interact with each other to form a 6-helical bundle – essential for viral and cell membrane fusion – several authors have reported HR1- and HR2-derived peptides that can inhibit this fusion [66–71]. Specifically, the fusion inhibitors HR2P, EK1, and EK1C4, exhibited a broad inhibitory activity against the SARS-CoV-2 and other SARSr- and MERSr-CoVs [66–71].

Additionally to the T-RHS and CDR described in the literature, we identified new potential hot spot residues which, to the best of our knowledge, have not been described before, regarding its druggability, structural importance and/or individual role in SARS-CoV-2. These include 181 (66%, 181/273) and 72 (65%, 72/110) residues identified in the S trimer structure of hSARSr-CoVs and SARSr- and MERS-CoVs, respectively. In both groups, these new potential hot spots lie essentially at the S2 subunit of the protein, particularly at the S2-NTD and CH domains (residues listed in Table S-7).

The potential hot spots residues are mostly surface exposed, within pockets of large volume and size, high enclosure and depth. They may represent advantageous targets for molecular and pharmacological modulation, since they potentially establish key interactions with host receptors or other molecules, or might play other roles in receptor recognition, S trimerization, S processing or in the mechanism of RBD conformational change. Moreover, the S protein represents a target for antibody-mediated neutralization by the host immune response and, consequently, constitutes a major antigenic component for structure-based vaccine design.

In this context, it is important to consider the effect and implications of the glycan shielding on immune evasion. Viral glycosylation not only mediates molecular recognition events, protein stability and folding, but it also plays a role in masking immunogenic protein epitopes from the host humoral immune system [72,73]. This could, in turn, hamper the neutralizing antibodies development. Several authors have recently predicted the SARS-CoV-2 S glycan shield pattern [52,72–74]. Casalino et al. characterized the glycosylation profile of the S protein and found that the HR2, TM and CT domains may be inaccessible to large molecules such as antibodies [72]. In this context, only small molecules should be explored when considering future antiviral strategies targeting these regions. On the contrary, the S head portion (including the S1-RBD, -SD1; SD2-FP, -CR, -HR1 and -CH domains, which are highlighted in our study) represents a more promising target, either for small molecules or for larger peptide and antibody-size molecules, since it is less shielded by the glycans [72]. Additionally, other authors suggest that the SARS-CoV-2 S protein is less densely glycosylated and, consequently, with a more vulnerable glycan shield when compared to other viral glycoproteins (HIV-1 envelope protein, Lassa virus glycoprotein complex and the influenza hemagglutinin) [73,74]. This may represent advantageous features in order to elicit SARS-CoV-2-neutralizing antibodies against the hot spots identified in our study. These authors have also identified glycosylation residues (N165 and N234) that may play a structural role in stabilizing the S1-RBD in the up conformation (determining the S open state) [72,73]; and the residue N234 represents a potential hot spot for drug targeting identified in our study.

The comprehensively structural characterization performed in our study should prompt the application of rational structure-based virtual screening, molecular docking and other *in silico*-chemico-biological approaches for the identification of potential novel CoV chemical inhibitors or monoclonal antibodies/vaccines.

Although the computational strategies have the potential to accelerate the drug discovery process and guide the posterior experimental approaches, they have limitations considering the experimental knowledge gaps [75,76]. Therefore, these potential hot spots should be experimentally studied *in vitro* and *in vivo* regarding its individual role in protein function or structure among beta-CoVs circulating in the human population. These data may also endorse the evaluation of SARS-CoV-2 S mutations resulting from the evolutionary adaptation of the virus to the human host.

The most relevant hot spot residues can be explored in discovery, design or development of chemical compounds or pan-Beta-CoV monoclonal antibodies that have the potential to inhibit the predicted pockets. Priority should be given to hot spots residues allocated to one or multiple consensus druggable pockets within the most promising S domains (and common to all S states). In this vein, a molecule designed to target one of these CDP could potentially interact with multiple sites/residues within the same CDP, or inhibit sites/residues that integrate multiple CDP within a major pocket. Consequently, other residues belonging to the considered CDP could be further studied in order to optimize or potentiate additional inhibitor-pocket binding interactions. A multi-target inhibitor can be, theoretically, more effective and less vulnerable to resistance. Hence, this rationale may also contribute to design inhibitors with a higher resilience to resistance since multiple mutations (in sites which have shown, *a priori*, high degree of conservation) would be required for the virus become resistant.

3. Conclusion

This study discloses a promising anti-CoV strategy directed to highly conserved druggable S regions, resulting from a comprehensive large-scale sequence analysis and structural characterization of protein domains across SARSr- and MERSr-CoVs.

The most promising conserved druggable regions highlighted in this study consisted of the RBD and SD1 (in the S1 subunit) and the FP, CR, HR1 and CH (in the S2 subunit) of S protein. Ultimately, the identification of CDR and conserved druggable sites/domains depends on the range of the anti-CoV strategy. A strategy which only targets the hSARSr-CoVs includes additional CDR within the RBD and FP domain, in addition to the CDR described for the RBD and SD1 (S1 subunit) of the SARSr- and MERSr-CoVs. Priority should be given to the RBD and SD1 (in S1) and to the CH (in S2), which exhibited a higher antiviral target potential among all protein conformations from all Beta-CoVs included in this study.

We have identified 181 new potential hot spot residues for hSARSr-CoVs and 72 new hot spot residues for SARSr- and MERSr-CoVs, which have not been described before in the literature. These hotspots are mostly surface exposed and represent attractive targets for molecular or pharmacological modulation. We hypothesize that an anti-CoV strategy targeting potential functionally or structurally highly conserved sites, provides a higher resilience to resistance development and can be potentially useful against the new SARS-CoV-2 and a broad spectrum of Beta-CoVs.

Further structural studies regarding the conservation and druggability of CoV proteins should also be extended to the RNA-dependent RNA-polymerase or to the main proteinase, considering its important function during the virus replication cycle. The strategies used in our study have been previously applied with success to structural and nonstructural proteins from other respiratory virus, and can be broadened applied to other re-emerging viral pathogens such as Zika virus, Ebola virus, as well as re-emerging CoV for a structure-based prediction of druggable biologic targets.

Ultimately, this study lays the foundation for successful structure-based design and discovery of chemical inhibitors, antibodies or other therapeutic modalities that target the S protein of Beta-CoVs.

4. Material and methods

4.1. Dataset construction and sequence analysis

All nucleotide sequences with a complete coding region of worldwide circulating human beta-CoVs including SARS-CoV-2, SARS-CoVs and MERS-CoVs, along with bat-SL-CoVs were retrieved on 24 March 2020 from sequence databases. These included: GISAID's EpiCov™ database (www.gisaid.org) [77,78], GenBank (<https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/>) [79], 2019 Novel Coronavirus Resource (2019nCoV) (<https://bigd.big.ac.cn/ncov>) [80], NIAID Virus Pathogen Resource (ViPR) database (<https://www.viprbrc.org>) [81] and NCBI Virus database (<https://www.ncbi.nlm.nih.gov/labs/virus>) [82].

The S surface glycoprotein encoded by the S gene was selected for study. The S gene was divided into the S1 and S2 subunits according to the S1/S2 cleavage site. The nucleotide sequence alignment was performed using Multiple Alignment Fast Fourier Transform (MAFFT) available at <https://www.ebi.ac.uk/Tools/msa/mafft/> [83,84]. The aa numbering was established according to the consensus sequence of the SARS-CoV-2 sequence alignment.

A total of 1,065 CoV S1 sequences and 1,096 CoV S2 sequences were included. Particularly, 674 SARS-CoV-2 isolates, 114 SARS-CoV isolates, 50 Bat-SL-CoV isolates and 248 MERS-CoV isolates for the S1 subunit, and 682 SARS-CoV-2 isolates, 116 SARS-CoV isolates, 50 Bat-SL-CoV isolates and 248 MERS-CoV isolates for the S2 subunit (depicted in Table 1).

All continents are represented with strains from the SARS-CoV-2. Specifically, Asian countries (n = 230 isolates) included Cambodia, China, Georgia, Hong Kong, India, Japan, Kuwait, Malaysia, Nepal, Saudi Arabia, Singapore, South Korea, Taiwan, Thailand and Vietnam. African countries (n = 3 isolates) included Democratic Republic of the Congo, Nigeria and South Africa. The North, Central and South America are represented by Canada, United States of America, Mexico, Panama, Brazil and Chile (n = 129 isolates). Twenty European countries were included in the study, namely Belgium, Czech Republic, Denmark, England, Finland, France, Germany, Hungary, Ireland, Italy, Lithuania, Luxemburg, Netherlands, Poland, Portugal, Scotland, Spain, Sweden, Switzerland and Wales (n = 298 isolates). Strains from Australia and New Zealand (Oceania) were also included (n = 26 isolates).

Multiple sequence alignment (MSA) visualization and the study of protein primary structure and length variation distributions among each beta-CoVs, was performed using MEGA X (<https://www.megasoftware.net/>) [85].

4.2. Conservation analysis

The aa conservation in each position was calculated using the Valdar scoring method from the Jalview AACons Web server (v2.11) [86,87]. The weighted scores incorporated into this conservation method consider the sequence redundancy in the MSA. This method enables the normalization against redundancy and bias in the MSA (reducing both the effect of bias sampling and penalizing gaps) with the benefit of no loss of evolutionary information [86]. This scoring system has been previously applied with success [88–90]. The conservation degree is presented as a numerical score, within a range of 0 (highly variable site) to 11 (highly conserved site), for each aa in the protein sequence alignment [86,87]. Residues with conservation score of 11 correspond to absolute conserved sites (100% aa identity). Residues were considered highly conserved for conservation score values ≥ 10 , conserved for conservation score of 7–9 and variable for conservation score values < 7 [90,91].

The SARS-CoV-2 was used as a reference for the alignment consensus (Jalview v 2.11.0 software). The conservation scores were calculated for each aa position within the S protein for three different groups: (a) SARS-CoV-2; (b) hSARSr-CoV; and (c) SARSr- and MERSr-CoV. A group comparing hSARS-CoVs and Bat-SL-CoVs was also briefly studied. The conservation scores for the (b) hSARSr-CoVs and the (c) SARSr- and MERSr-CoVs were mapped onto the S monomeric structure (SARS-CoV-2, PDB entry 6VYB).

4.3. Druggability analysis

Druggability studies were performed using the newly available SARS-CoV-2 crystallographic structures. Crystallographic structures of SARS-CoV-2 S protein trimer (PDB entries 6VSB, 6VXX, 6VYB) [6,92] and RBD (PDB entries 6VW1, 6LZG, 6M0J) [12,28,44] were selected for study and retrieved from the RCSB Protein Data Bank (www.rcsb.org) [93]. Each structure was accurately prepared by removing coordinated molecules and bounded ligands, protonation and energy minimization using MOE 2015.10001 program [94]. To date, the available S crystal structures are not resolved for the signal peptide (S1 subunit), heptad repeat 2 (HR2), transmembrane region and CP (S2 subunit), which limit the druggability prediction and further analysis of these regions.

The druggable sites/residues within the selected structures were predicted from a consensus of a triple strategy, following the same rational described by Trigueiro-Louro et al. 2019, using the webserver DGSS at <https://proteins.plus>, the commercial software MOE-SiteFinder (SF) from Chemical Computing Group and (when applied) the PockDrug-Server (PDS) at <http://pockdrug.rpbs.univ-paris-diderot.fr/> [90,94–97].

The DGSS server is based on a grid-based method incorporated in a support vector machine model for druggability predictions. The algorithm is based on geometric and physicochemical descriptors and provides a raw druggability score for each pocket, ranging from 0 (undruggable) to 1 (druggable). Pockets with a DGSS druggability score ≥ 0.4 were considered druggable [96]. The druggability threshold was previously defined by Volkamer and the methodology was applied with success by other authors [90,96,98–100].

A triple consensus-based strategy was applied since some variation may be observed among different druggability prediction methods [90,97]. Considering the PockDrug-Server is able to predict the pocket(s) druggability for a maximum file size of 1 MB, only the crystallographic structures of the SARS-Cov-2 S RBD were studied using this webserver.

The druggability prediction for the S monomer and trimer structures (PDB IDs: 6VSB, 6VXX, 6VYB) was studied for each individual structure using the bioinformatics tools: (a) SF and (b) DGSS and according to the number of structures ascribed to each group. The druggability analysis for each individual SARS-CoV-2 S RBD structure (PDB entries 6VW1, 6LZG, 6M0J) was performed with the three bioinformatic tools: (a) SF, (b) DGSS and (c) PDS. Subsequently, a comparative analysis was performed by merging the druggability information from the three bioinformatic tools for each group: (1) S monomer structures; (2) S RBD structures; and (3) S trimer structures.

The last stage of the study consisted of a global comparative analysis of the most prevailing druggable sites identified by the bioinformatics tools and shared by all the structures within each group. The druggability consensus of each group was manually analysed in parallel with the conservation data to identify druggable binding sites/pockets or potential hot spots residues that overlap to the conserved regions/residues of the CoV S proteins.

The coordinates of the highly score CDPs were mapped onto the three-dimensional S trimer structure using PyMOL Molecular Graphics System, version 1.8 (Schrödinger, LLC; <https://pymol.org>).

[101]. This rational has been previously applied with success in studies performed by the same authors and other studies regarding the druggability analysis of influenza virus proteins [90,102,103].

CRedit authorship contribution statement

João Trigueiro-Louro: Conceptualization, Methodology, Software, Investigation, Writing - original draft, Writing - review & editing, Supervision. **Vanessa Correia:** Methodology, Software, Investigation, Writing - original draft, Writing - review & editing. **Inês Figueiredo-Nunes:** Methodology, Software. **Marta Gíria:** Investigation, Writing - original draft, Writing - review & editing. **Helena Rebelo-de-Andrade:** Conceptualization, Methodology, Investigation, Writing - review & editing, Supervision, Project administration, Funding acquisition.

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

The authors would like to thank to all authors, originating and submitting laboratories that share their sequences through the GISAID EpiCoV™, GenBank, 2019nCoV, ViPR and NCBI Virus database.

The authors thank Rita Guedes Lab Group (Medicinal Chemistry Unit, iMed.Ulisboa, FFUL), namely Rita Guedes, Patrícia Serra and Rita Acúrcio for providing the access to computational equipment required for this study and for sharing their expertise.

The authors also thank to Maria de São José Alexandre (Universidade do Porto) and Rui M.M. Brito (Centro de Química de Coimbra, Universidade de Coimbra) for the guidance given as supervisors of the Ph.D. work of J.Trigueiro-Louro.

Funding

This work is funded by National Funds through the FCT - Fundação para a Ciência e a Tecnologia, I.P., by the FCT project PTDC/SAU-INF/30729/2017; and supported by the PhD grant PD/BD/128402/2017 from FCT PhD Programme in Medicines and Pharmaceutical Innovation (i3DU).

FCT
Fundação
para a Ciência
e a Tecnologia



REPÚBLICA PORTUGUESA

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2020.07.017>.

References

- [1] NIH. COVID-19 Treatment Guidelines Panel. Coronavirus Disease 2019 (COVID-19) Treatment Guidelines. National Institutes of Health 2020. <https://www.covid19treatmentguidelines.nih.gov/> [accessed 29 May 2020].
- [2] Abd El-Aziz TM, Stockand JD. Recent progress and challenges in drug development against COVID-19 coronavirus (SARS-CoV-2) - an update on the status. *Infect Genet Evol* 2020. <https://doi.org/10.1016/j.meegid.2020.104327>.
- [3] Li F. Structure, function, and evolution of coronavirus spike proteins. *Annu Rev Virol* 2016. <https://doi.org/10.1146/annurev-virology-110615-042301>.
- [4] 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* 2020. <https://doi.org/10.1016/j.cell.2020.02.052>.
- [5] Zhou P, Lou YX, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020. <https://doi.org/10.1038/s41586-020-2012-7>.
- [6] Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 2020. <https://doi.org/10.1016/j.cell.2020.02.058>.
- [7] Zumla A, Chan JFW, Azhar EI, Hui DSC, Yuen KY. Coronaviruses—drug discovery and therapeutic options. *Nat Rev Drug Discov* 2016. <https://doi.org/10.1038/nrd.2015.37>.
- [8] Menachery VD, Dinnon KH, Yount BL, McAnarney ET, Gralinski LE, Hale A, et al. Trypsin treatment unlocks barrier for zoonotic bat coronavirus infection. *J Virol* 2019. <https://doi.org/10.1128/jvi.01774-19>.
- [9] Yang Y, Du L, Liu C, Wang L, Ma C, Tang J, et al. Receptor usage and cell entry of bat coronavirus HKU4 provide insight into bat-to-human transmission of MERS coronavirus. *Proc Natl Acad Sci U S A* 2014. <https://doi.org/10.1073/pnas.1405889111>.
- [10] Yang Y, Liu C, Du L, Jiang S, Shi Z, Baric RS, et al. Two mutations were critical for bat-to-human transmission of middle east respiratory syndrome coronavirus. *J Virol* 2015. <https://doi.org/10.1128/jvi.01279-15>.
- [11] Hoffmann M, Hofmann-Winkler H, Pöhlmann S. Priming time: how cellular proteases arm coronavirus spike proteins. *Act. Viruses by Host Proteases* 2018. https://doi.org/10.1007/978-3-319-75474-1_4h.
- [12] Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 2020. <https://doi.org/10.1038/s41586-020-2180-5>.
- [13] Ou X, Liu Y, Lei X, Li P, Mi D, Ren L, et al. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat Commun* 2020. <https://doi.org/10.1038/s41467-020-15562-9>.
- [14] Gierer S, Bertram S, Kaup F, Wrensch F, Heurich A, Kramer-Kuhl A, et al. The spike protein of the emerging betacoronavirus EMC uses a novel coronavirus receptor for entry, can be activated by TMPRSS2, and is targeted by neutralizing antibodies. *J Virol* 2013. <https://doi.org/10.1128/jvi.00128-13>.
- [15] Glowacka I, Bertram S, Muller MA, Allen P, Soilleux E, Pfefferle S, et al. Evidence that TMPRSS2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response. *J Virol* 2011. <https://doi.org/10.1128/jvi.02232-10>.
- [16] Iwata-Yoshikawa N, Okamura T, Shimizu Y, Hasegawa H, Takeda M, Nagata N. TMPRSS2 contributes to virus spread and immunopathology in the airways of murine models after coronavirus infection. *J Virol* 2019. <https://doi.org/10.1128/jvi.01815-18>.
- [17] Matsuyama S, Nagata N, Shirato K, Kawase M, Takeda M, Taguchi F. Efficient activation of the severe acute respiratory syndrome coronavirus spike protein by the transmembrane protease TMPRSS2. *J Virol* 2010. <https://doi.org/10.1128/jvi.01542-10>.
- [18] Shulla A, Heald-Sargent T, Subramanya G, Zhao J, Perlman S, Gallagher T. A transmembrane serine protease is linked to the severe acute respiratory syndrome coronavirus receptor and activates virus entry. *J Virol* 2011. <https://doi.org/10.1128/jvi.02062-10>.
- [19] Zhou Y, Vedantham P, Lu K, Agudelo J, Carrion R, Nunneley JW, et al. Protease inhibitors targeting coronavirus and filovirus entry. *Antiviral Res* 2015. <https://doi.org/10.1016/j.antiviral.2015.01.011>.
- [20] Kirchmair J, Distinto S, Liedl KR, Markt P, Rollinger JM, Schuster D, et al. Development of anti-viral agents using molecular modeling and virtual screening techniques. *Infect Disord Drug Targets* 2011;11:64–93.
- [21] Mallipeddi PL, Kumar G, White SW, Webb TR. Recent advances in computer-aided drug design as applied to anti-influenza drug discovery. *Curr Top Med Chem* 2014;14:1875–89.
- [22] Mohs RC, Greig NH. Drug discovery and development: Role of basic biological research. *Alzheimers Dement (N Y)* 2017;3:651–7. <https://doi.org/10.1016/j.trci.2017.10.005>.
- [23] Schuhmacher A, Gassmann O, Hinder M. Changing R&D models in research-based pharmaceutical companies. *J Transl Med* 2016;14:105. <https://doi.org/10.1186/s12967-016-0838-4>.
- [24] Ng OW, Tan YJ. Understanding bat SARS-like coronaviruses for the preparation of future coronavirus outbreaks – implications for coronavirus vaccine development. *Hum Vaccines Immunother* 2017. <https://doi.org/10.1080/21645515.2016.1228500>.
- [25] He B, Zhang Y, Xu L, Yang W, Yang F, Feng Y, et al. Identification of diverse alphacoronaviruses and genomic characterization of a novel severe acute respiratory syndrome-like coronavirus from bats in China. *J Virol* 2014. <https://doi.org/10.1128/jvi.00631-14>.
- [26] Kleine-Weber H, Elzayat MT, Hoffmann M, Pöhlmann S. Functional analysis of potential cleavage sites in the MERS-coronavirus spike protein. *Sci Rep* 2018. <https://doi.org/10.1038/s41598-018-34859-w>.
- [27] Gierer S, Müller MA, Heurich A, Ritz D, Springstein BL, Karsten CB, et al. Inhibition of proprotein convertases abrogates processing of the middle eastern respiratory syndrome coronavirus spike protein in infected cells but does not reduce viral infectivity. *J Infect Dis* 2015. <https://doi.org/10.1093/infdis/jiu407>.
- [28] Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, et al. Structural basis of receptor recognition by SARS-CoV-2. *Nature* 2020. <https://doi.org/10.1038/s41586-020-2179-y>.
- [29] Du L, He Y, Zhou Y, Liu S, Zheng BJ, Jiang S. The spike protein of SARS-CoV - a target for vaccine and therapeutic development. *Nat Rev Microbiol* 2009. <https://doi.org/10.1038/nrmicro2090>.
- [30] Du L, Yang Y, Zhou Y, Lu L, Li F, Jiang S. MERS-CoV spike protein: a key target for antivirals. *Expert Opin Ther Targets* 2017. <https://doi.org/10.1080/14728222.2017.1271415>.
- [31] Prajapat M, Sarma P, Shekhar N, Avti P, Sinha S, Kaur H, et al. Drug targets for corona virus: a systematic review. *Indian J Pharmacol* 2020. https://doi.org/10.4103/ijp.IJP_115_20.
- [32] Totura AL, Bavari S. Broad-spectrum coronavirus antiviral drug discovery. *Expert Opin Drug Discov* 2019. <https://doi.org/10.1080/17460441.2019.1581171>.
- [33] Hajduk PJ, Huth JR, Tse C. Predicting protein druggability. *Drug Discov Today* 2005;10:1675–82. [https://doi.org/10.1016/S1359-6446\(05\)03624-X](https://doi.org/10.1016/S1359-6446(05)03624-X).
- [34] Gao M, Skolnick J. APoc: large-scale identification of similar protein pockets. *Bioinformatics* 2013. <https://doi.org/10.1093/bioinformatics/btt024>.
- [35] Perola E, Herman L, Weiss J. Development of a rule-based method for the assessment of protein druggability. *J Chem Inf Model* 2012;52:1027–38. <https://doi.org/10.1021/ci200613b>.
- [36] Gur M, Taka E, Yilmaz SZ, Kilinc C, Aktas U, Exploring GM, et al. Novel coronavirus spike glycoprotein between its closed and open states using molecular dynamics simulations. *BioRxiv* 2019;2020. <https://doi.org/10.1101/2020.04.17.047324>.
- [37] Song W, Gui M, Wang X, Xiang Y. Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS Pathog* 2018. <https://doi.org/10.1371/journal.ppat.1007236>.
- [38] Walls AC, Tortorici MA, Snijder J, Xiong X, Bosch BJ, Rey FA, et al. Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proc Natl Acad Sci U S A* 2017. <https://doi.org/10.1073/pnas.1708727114>.
- [39] Li T, Zheng Q, Yu H, Wu D, Xue W, Zhang Y, et al. Characterization of the SARS-CoV-2 spike in an early prefusion conformation. *BioRxiv* 2020. <https://doi.org/10.1101/2020.03.16.994152>.
- [40] Walls A, Tortorici MA, Bosch BJ, Frenz B, Rottier PJM, DiMaio F, et al. Crucial steps in the structure determination of a coronavirus spike glycoprotein using cryo-electron microscopy. *Protein Sci* 2017. <https://doi.org/10.1002/pro.3048>.
- [41] Kim SY, Jin W, Sood A, Montgomery DW, Grant OC, Fuster MM, et al. Glycosaminoglycan binding motif at S1/S2 proteolytic cleavage site on spike glycoprotein may facilitate novel coronavirus (SARS-CoV-2) host cell entry. *BioRxiv* 2020.04.14.041459, doi: 10.1101/2020.04.14.041459.
- [42] Li F, Berardi M, Li W, Farzan M, Dormitzer PR, Harrison SC. Conformational states of the severe acute respiratory syndrome coronavirus spike protein ectodomain. *J Virol* 2006. <https://doi.org/10.1128/jvi.02744-05>.
- [43] Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 2014. <https://doi.org/10.1093/nar/gku316>.
- [44] Wang Q, Zhang Y, Wu L, Niu S, Song C, Zhang Z, et al. Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell* 2020. <https://doi.org/10.1016/j.cell.2020.03.045>.
- [45] Qing E, Hantak M, Perlman S, Gallagher T. Distinct roles for sialoside and protein receptors in coronavirus infection. *MBio* 2020. <https://doi.org/10.1128/mBio.02764-19>.
- [46] Wang C, Li W, Drabek D, Okba NMA, van Haperen R, Osterhaus ADME, et al. A human monoclonal 1 antibody blocking SARS-CoV-2 infection. *BioRxiv* 2020. <https://doi.org/10.1101/2020.03.11.987958>.

- [47] Pinto D, Park Y-J, Beltramello M, Walls AC, Tortorici MA, Bianchi S, et al. Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature* 2020. <https://doi.org/10.1038/s41586-020-2349-y>.
- [48] Li L, Sun T, He Y, Li W, Fan Y, Zhang J. Epitope-based peptide vaccine design and target site characterization against novel coronavirus disease caused by SARS-CoV-2. *BioRxiv* 2020:2020.02.25.965434, doi: 10.1101/2020.02.25.965434.
- [49] Grifoni A, Sidney J, Zhang Y, Scheuermann RH, Peters B, Sette A. A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. *Cell Host Microbe* 2020. <https://doi.org/10.1016/j.chom.2020.03.002>.
- [50] Aydin H, Al-Khooly D, Lee JE. Influence of hydrophobic and electrostatic residues on SARS-coronavirus S2 protein stability: insights into mechanisms of general viral fusion and inhibitor design. *Protein Sci* 2014. <https://doi.org/10.1002/pro.2442>.
- [51] Robson B. Computers and viral diseases. Preliminary bioinformatics studies on the design of a synthetic vaccine and a preventative peptidomimetic antagonist against the SARS-CoV-2 (2019-nCoV, COVID-19) coronavirus. *Comput Biol Med* 2020. <https://doi.org/10.1016/j.compbio.2020.103670>.
- [52] Vankadari N, Wilce JA. Emerging WuHan (COVID-19) coronavirus: glycan shield and structure prediction of spike glycoprotein and its interaction with human CD26. *Emerg Microbes Infect* 2020. <https://doi.org/10.1080/22221751.2020.1739565>.
- [53] Bongini P, Trezza A, Bianchini M, Spiga O, Niccolai N. A possible strategy to fight COVID-19: interfering with spike glycoprotein trimerization. *Biochem Biophys Res Commun* 2020. <https://doi.org/10.1016/j.bbrc.2020.04.007>.
- [54] Kalathiya U, Padariya M, Mayordomo M, Lisowska M, Nicholson J, Singh A, et al. Highly conserved homotrimer cavity formed by the SARS-CoV-2 spike glycoprotein: a novel binding site. *J Clin Med* 2020. <https://doi.org/10.3390/jcm9051473>.
- [55] Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* 2020. <https://doi.org/10.1016/j.cell.2020.06.043>.
- [56] Chen SH, Young MT, Gounley J, Stanley C, Bhowmik D. Distinct structural flexibility within SARS-CoV-2 spike protein reveals potential therapeutic targets. *BioRxiv* 2020. <https://doi.org/10.1101/2020.04.17.047548>.
- [57] Peters MH, Bastidas O. Static All-Atom Energetic Mappings of the SARS-Cov-2 Spike Protein with Potential Latch Identification of the Down State Protomer. 2020, doi: 10.1101/2020.05.12.091090.
- [58] Li Z, Tomlinson ACA, Wong AHM, Zhou D, Desforges M, Talbot PJ, et al. The human coronavirus HCoV-229E S-protein structure and receptor binding. *Elife* 2019. <https://doi.org/10.7554/elife.51230>.
- [59] Choudhury A, Mukherjee S. In silico studies on the comparative characterization of the interactions of SARS-CoV-2 spike glycoprotein with ACE-2 receptor homologs and human TLRs. *J Med Virol* 2020. <https://doi.org/10.1002/jmv.25987>.
- [60] Fantini J, Di Scala C, Chahinian H, Yahi N. Structural and molecular modelling studies reveal a new mechanism of action of chloroquine and hydroxychloroquine against SARS-CoV-2 infection. *Int J Antimicrob Agents* 2020. <https://doi.org/10.1016/j.ijantimicag.2020.105960>.
- [61] Fantini J, Chahinian H, Yahi N. Synergistic antiviral effect of hydroxychloroquine and azithromycin in combination against SARS-CoV-2: What molecular dynamics studies of virus-host interactions reveal. *Int J Antimicrob Agents* 2020;106020.
- [62] Hodos RA, Kidd BA, Shameer K, Readhead BP, Dudley JT. In silico methods for drug repurposing and pharmacology. *Wiley Interdiscip Rev Syst Biol Med* 2016. <https://doi.org/10.1002/wsbm.1337>.
- [63] Vankadari N. Arbidol: A potential antiviral drug for the treatment of SARS-CoV-2 by blocking the trimerization of viral spike glycoprotein? *Int J Antimicrob Agents* 2020, doi: 10.1016/j.ijantimicag.2020.105998.
- [64] Musarrat F, Chouljenko V, Dahal A, Nabi R, Chouljenko T, Jois SD, et al. The anti-HIV drug nelfinavir mesylate (Viracept) is a potent inhibitor of cell fusion caused by the SARS-CoV-2 spike (S) glycoprotein warranting further evaluation as an antiviral against COVID-19 infections. *J Med Virol* 2020. <https://doi.org/10.1002/jmv.25985>.
- [65] Salman S, Shah FH, Idrees J, Idrees F, Velagala S, Ali J, et al. Virtual screening of immunomodulatory medicinal compounds as promising anti-SARS-CoV-2 inhibitors. *Future Virol* 2020. <https://doi.org/10.2217/fvl-2020-0079>.
- [66] Liu S, Xiao G, Chen Y, He Y, Niu J, Escalante CR, et al. Interaction between heptad repeat 1 and 2 regions in spike protein of SARS-associated coronavirus: implications for virus fusogenic mechanism and identification of fusion inhibitors. *Lancet* 2004. [https://doi.org/10.1016/S0140-6736\(04\)15788-7](https://doi.org/10.1016/S0140-6736(04)15788-7).
- [67] Lu L, Liu Q, Zhu Y, Chan KH, Qin L, Li Y, et al. Structure-based discovery of Middle East respiratory syndrome coronavirus fusion inhibitor. *Nat Commun* 2014. <https://doi.org/10.1038/ncomms4067>.
- [68] Xia S, Lan Q, Pu J, Wang C, Liu Z, Xu W, et al. Potent MERS-CoV fusion inhibitory peptides identified from HR2 domain in spike protein of bat coronavirus HKU4. *Viruses* 2019. <https://doi.org/10.3390/v11010056>.
- [69] Xia S, Yan L, Xu W, Agrawal AS, Algaissi A, Tseng CTK, et al. A pan-coronavirus fusion inhibitor targeting the HR1 domain of human coronavirus spike. *Sci Adv* 2019. <https://doi.org/10.1126/sciadv.aav4580>.
- [70] Xia S, Zhu Y, Liu M, Lan Q, Xu W, Wu Y, et al. Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. *Cell Mol Immunol* 2020. <https://doi.org/10.1038/s41423-020-0374-2>.
- [71] Xia S, Liu M, Wang C, Xu W, Lan Q, Feng S, et al. Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. *Cell Res* 2020. <https://doi.org/10.1038/s41422-020-0305-x>.
- [72] Casalino L, Gaieb Z, Dommer AC, Harbison AM, Fogarty CA, Barros EP, et al. Shielding and beyond: the roles of glycans in SARS-CoV-2 spike protein. *BioRxiv* 2020. <https://doi.org/10.1101/2020.06.11.146522>.
- [73] Watanabe Y, Allen JD, Wrapp D, McLellan JS, Crispin M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* (80-) 2020. <https://doi.org/10.1126/science.abb9983>.
- [74] Watanabe Y, Berndsen ZT, Raghwanji J, Seabright GE, Allen JD, Pybus OG, et al. Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. *Nat Commun* 2020. <https://doi.org/10.1038/s41467-020-16567-0>.
- [75] Sliwoski G, Kothiwale S, Meiler J, Lowe Jr EW. Computational methods in drug discovery. *Pharmacol Rev* 2014;66:334–95. <https://doi.org/10.1124/pr.112.007336>.
- [76] Cichonska A, Ravikumar B, Parri E, Timonen S, Pahikkala T, Airola A, et al. Computational-experimental approach to drug-target interaction mapping: A case study on kinase inhibitors e1005678 e1005678. *PLoS Comput Biol* 2017;13. <https://doi.org/10.1371/journal.pcbi.1005678>.
- [77] Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Challenges* 2017. <https://doi.org/10.1002/gch2.1018>.
- [78] Shu Y, McCauley J. GISAID: global initiative on sharing all influenza data - from vision to reality. *Euro Surveill* 2017;22. <https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494>.
- [79] Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Ostell J, Pruitt KD, et al. GenBank. *Nucleic Acids Res* 2018. <https://doi.org/10.1093/nar/gkx1094>.
- [80] Zhao W-M, Song S-H, Chen M-L, Zou D, Ma L-N, Ma Y-K, et al. The 2019 novel coronavirus resource. *Yi Chuan = Hered* 2020.
- [81] Pickett BE, Sadat EL, Zhang Y, Noronha JM, Squires RB, Hunt V, et al. ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic Acids Res* 2012. <https://doi.org/10.1093/nar/gkr859>.
- [82] Hatcher EL, Zhdanov SA, Bao Y, Blinkova O, Nawrocki EP, Ostapchuk Y, et al. Virus variation resource-improved response to emergent viral outbreaks. *Nucleic Acids Res* 2017. <https://doi.org/10.1093/nar/gkw1065>.
- [83] Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* 2019. <https://doi.org/10.1093/nar/gkz268>.
- [84] Katoh K. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002. <https://doi.org/10.1093/nar/gkf436>.
- [85] Kumar S, Stecher G, Li M, Nnyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 2018. <https://doi.org/10.1093/molbev/msy096>.
- [86] Valdar WS. Scoring residue conservation. *Proteins* 2002;48:227–41. <https://doi.org/10.1002/prot.10146>.
- [87] Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 2009;25:1189–91. <https://doi.org/10.1093/bioinformatics/btp033>.
- [88] Patel H, Kukol A. Evolutionary conservation of influenza A PB2 sequences reveals potential target sites for small molecule inhibitors. *Virology* 2017. <https://doi.org/10.1016/j.virol.2017.06.009>.
- [89] Kukol A. *Combining biology and chemistry to target the flu virus*. Beau Bassin: Lap Lambert Academic Publishing; 2017.
- [90] Trigueiro-Louro JM, Correia V, Santos LA, Guedes RC, Brito RMM, Rebelo-Andrade H. To hit or not to hit: large-scale sequence analysis and structure characterization of influenza A NS1 unlocks new antiviral target potential. *Virology* 2019. <https://doi.org/10.1016/j.virol.2019.04.009>.
- [91] Livingstone CD, Barton GJ. Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Comput Appl Biosci* 1993;9:745–56.
- [92] Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* (80-) 2020. <https://doi.org/10.1126/science.aaa0902>.
- [93] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res* 2000;28:235–42.
- [94] Molecular Operating Environment (MOE) 2013.08. Molecular Operating Environment (MOE), 2013.08; Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7. Mol Oper Environ (MOE), 201308; Chem Comput Gr Inc, 1010 Sherbooke St West, Suite #910, Montr QC, Canada, H3A 2R7, 2013 2016.
- [95] Hussein HA, Borrel A, Geneix C, Petitjean M, Regad L, Camproux AC. PockDrug-Server: a new web server for predicting pocket druggability on holo and apo proteins. *Nucleic Acids Res* 2015;43:W436–42. <https://doi.org/10.1093/nar/gkv462>.
- [96] Volkamer A, Kuhn D, Rippmann F, Rarey M. DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics* 2012;28:2074–5. <https://doi.org/10.1093/bioinformatics/bts310>.
- [97] Volkamer A, Griewel A, Grombacher T, Rarey M. Analyzing the topology of active sites: on the prediction of pockets and subpockets. *J Chem Inf Model* 2010. <https://doi.org/10.1021/ci100241v>.

- [98] Correia V, Abecasis AB, Rebelo-de-Andrade H. Molecular footprints of selective pressure in the neuraminidase gene of currently circulating human influenza subtypes and lineages. *Virology* 2018. <https://doi.org/10.1016/j.virol.2018.07.002>.
- [99] Ehart C, Brinkjost T, Koch O. Binding site characterization-similarity, promiscuity, and druggability. *Medchemcomm* 2019. <https://doi.org/10.1039/c9md00102f>.
- [100] Irsheid L, Wehler T, Borek C, Kiefer W, Brenk R, Ortiz-Soto ME, et al. Identification of a potential allosteric site of Golgi α -mannosidase II using computer-aided drug design. *PLoS ONE* 2019. <https://doi.org/10.1371/journal.pone.0216132>.
- [101] Schrodinger LLC. The PyMOL Molecular Graphics System. Version 2015;1:8.
- [102] Kukol A, Hughes DJ. Large-scale analysis of influenza A virus nucleoprotein sequence conservation reveals potential drug-target sites. *Virology* 2014;454–455:40–7. <https://doi.org/10.1016/j.virol.2014.01.023>.
- [103] Darapaneni V, Prabhaker VK, Kukol A. Large-scale analysis of influenza A virus sequences reveals potential drug target sites of non-structural proteins. *J Gen Virol* 2009;90:2124–33. <https://doi.org/10.1099/vir.0.011270-0>.