Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Structural insights of key enzymes into therapeutic intervention against SARS-CoV-2

Munazza Shahid, Syed Shahzad-ul-Hussan *

Department of Biology, SBA School of Science and Engineering, Lahore University of Management Sciences, Lahore 54792, Pakistan

ARTICLE INFO
Keywords: SARS-CoV-2 COVID-19 Polymerase 3CLpro Therapeutic targets α-Glucosidase inhibitor

ABSTRACT
COVID-19 pandemic, caused by SARS-CoV-2, has drastically affected human health all over the world. After the emergence of the pandemic the major focus of efforts to attenuate the infection has been on repurposing the already approved drugs to treat COVID-19 adopting a fast-track strategy. However, to date a specific regimen to treat COVID-19 is not available. Over the last few months a substantial amount of data about the structures of various key proteins and their recognition partners involved in the SARS-CoV-2 pathogenesis has emerged. These studies have not only provided the molecular level descriptions of the viral pathogenesis but also laid the foundation for rational drug design and discovery. In this review, we have recapitulated the structural details of four key viral enzymes, RNA-dependent RNA polymerase, 3-chymotrypsin like protease, papain-like protease and helicase, and two host factors including angiotensin-converting enzyme 2 and transmembrane serine protease involved in the SARS-CoV-2 pathogenesis, and described the potential hotspots present on these structures which could be explored for therapeutic intervention. We have also discussed the significance of endoplasmic reticulum α-glucosidases as potential targets for anti-SARS-CoV-2 drug discovery.

1. Introduction

The current coronavirus infectious disease 19 (COVID-19) pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has shown unprecedented impact on healthcare, economies and social life all over the world (Bayham and Fenichel, 2020; Burkle, 2020). As of September 21st, 2020 over 30.9 million people have been confirmed with COVID-19 with over 959,000 deaths since the emergence of the infection (WHO, 2020). An effective vaccine to prevent the infection yet to be available despite a significant number of vaccine candidates have been reported which are in pre-clinical development or under clinical trials (Al-Kassmy et al., 2020; Checcucci et al., 2020; Wang et al., 2020). Any specific treatment regimen against COVID-19 is also not available to date.

SARS-CoV-2 belongs to the Coronavirinae subfamily in the Coronavirus family of the Nidovirales order (Lu et al., 2020). Angiotensin-converting enzyme 2 (ACE2) present at the cellular surface of various types of tissues acts as a primary receptor of SARS-CoV-2. The receptor-binding domain (RBD) in the S1 subunit of the viral spike (S) protein mediates the binding of the virus to ACE2. Subsequently, priming of the S protein by a host protease enzyme, transmembrane serine protease 2 (TMPRSS2) enables the fusion of membranes and viral entry (Hoffmann et al., 2020; Wan et al., 2020; Wrapp and Wang, 2020b).

SARS-CoV-2 is a positive sense RNA virus with genome size of 29.9 kb which contains 11 open reading frames (ORF). Orf1a and orf1ab encode the pp1a and pp1ab replicase polyproteins (Fig. 1) (Chan and Kok, 2020; Licastro et al., 2020). The replicase polyproteins are cleaved by papain like protease (PLpro) at 3 different N-terminus sites and by 3-chymotrypsin-like protease (3CLpro) at 11 different sites to release non-structural proteins. Overall, the SARS-CoV-2 genome encodes 16 non-structural proteins (nspl-nsp16), 4 structural proteins including spike, the membrane protein, envelope and nucleocapsid, and accessory proteins 3, 6, 7a, 7b, 8, and 10. Non-structural proteins are translated from viral genomic RNA in the form of replicase polyproteins while structural and accessory proteins are translated from sub-genomic RNAs. Six of the non-structural proteins are enzymes in nature, which include the PLpro domain in nsp3, nsp5 (3CLpro), nsp12 (RNA-dependent RNA polymerase), nsp13 (helicase), nsp14 (N7-methyltransferase) and nsp16 (2′-O-methyltransferase), while nsp7-nsp10 play regulatory roles in the replication cycle of the virus (Chellapandi and Saranya, 2020; Lei et al., 2020; Shereen et al., 2020).

After the emergence of the COVID-19 pandemic, early efforts to
discover treatments were focused on evaluating the efficacy of already known drugs particularly antiviral agents against SARS-CoV-2. Several clinically approved drugs including RdRp inhibitor remdesivir that was initially approved to treat Ebola virus, a well known antimalarial drug chloroquine and its derivative hydroxychloroquine and a combination therapy consisting of HIV protease inhibitors, ritonavir and lopinavir were approved by the World Health Organization (WHO) in March 2020 for the SOLIDARITY worldwide clinical trials, to evaluate the clinical efficacy of these drugs against COVID-19 (Cao et al., 2020; Gadebusch Bondio and Marloth, 2020; Hung et al., 2020; Sterling and Irwin, 2015; Wang et al., 2020b, 2020d). Drug repurposing and various clinical trials of already approved drugs to evaluate their efficacy against COVID-19 have been reviewed in detail by several authors (Lima et al., 2020; Pandey et al., 2020; Rahman et al., 2020; Santos et al., 2020; Tu et al., 2020).

Considering the global impact of the COVID-19 pandemic and possibility of re-emergence of coronavirus infections in future, there is an urgent need of developing new antiviral agents particularly with broad-spectrum efficacy against different coronaviruses. In this regard, different viral proteins involved in the viral replication process that are highly conserved or contain conserved motifs among coronaviruses, represent potential targets for broad-spectrum anti-coronavirus therapeutic development. Several recently published reviews have provided an overview of the coronavirus therapeutic targets and details of different in vitro and/or in vivo characterized anti-coronavirus active agents (McKee et al., 2020; Su et al., 2020a; Tiwari et al., 2020). Targeting specific structural features of the proteins crucial for viral replication cycle to interfere with their functions is an important strategy to identify new antiviral drugs.

Advancement in biotechnology to efficiently produce viral proteins and structure determination techniques particularly cryogenic electron microscopy (cryo-EM) along with high commitment of the scientific community during the pandemic have led the emergence of structures of various key proteins of SARS-CoV-2 also including novel scaffolds (Chen et al., 2020; Gao et al., 2020; Hillen and Kocic, 2020; Jin et al., 2020; Lan et al., 2020; Melero et al., 2020; Shang et al., 2020; Shin and Mukherjee, 2020; te Velthuis et al., 2012; Wrapp and Wang, 2020a; Yin et al., 2020). In this review we recapitulated the structural features of the key catalytic proteins involved in the SARS-CoV-2 pathogenesis, their main catalytic sites and other structural regions involved in interactions with partner proteins resulting in functional outcomes crucial for the viral pathogenesis and propagation. We have also highlighted the known inhibitors of these proteins in the context of describing their interactions with the target proteins according to the available structural data. Plenty of studies on different aspects of SARS-CoV-2 are emerging, however, here we focused on four viral enzymes, PLpro, 3CLpro, helicase and RdRp, and two host factors including ACE2, TMPRSS2 to discuss potential structural hotspots that could be explored as target sites for therapeutic intervention. Moreover, we also discussed α-glucosidases of human endoplasmic reticulum (ER) as potential therapeutic targets against SARS-CoV-2.

2. Key enzymes of SARS-CoV-2 replication cycle

2.1. PLpro of SARS-CoV-2

The PLpro enzyme of coronaviruses is a catalytic domain of the large multidomain nsp3 protein (Lei et al., 2018). It cleaves the pp1a/ab polyprotein at the junctions of nsp1/nsp2, nsp2/nsp3 and nsp3/nsp4 releasing nsp1, nsp2 and nsp3 thereby playing important role in the viral replication cycle. PLpro recognizes a conserved sequence consisting of Leu-X-Gly-Gly and cleaves at the carboxyl-terminal of the last glycine residue. SARS-CoV-2 PLpro shows 83% sequence identity with its counterpart of SARS-CoV.

To date several crystal structures of SCoV-2 PLpro and its complexes with different ligands have been reported in protein data banks. Structurally, the enzyme consists of a canonical right-handed thumb, palm and fingers architecture also containing a forth N-terminus ubiquitin-like sub-domain (Fig. 2a). The catalytic site of PLpro is conserved among all coronaviruses and consists of a catalytic triad of Cys-111, His-272 and Asp-286. During catalysis, Cys-111 acts as a nucleophile, His-272 plays a role of general acid/base and Asp-286 promotes the deprotonation of Cys-111. The catalytic site is present in a ridge between the palm and the thumb domains (Shin and Mukherjee, 2020).

Recently, Rut et al investigated the substrate specificity of PLpro by using a combinatorial library of tetrapeptides where systematic variations of natural and unnatural amino acids were incorporated at positions P2, P3 and P4 by keeping Gly at the P1 position. The study suggested very high specificity of the enzyme for Gly at the P2 site, preference for positive and hydrophobic residues at the P3 site and a high preference for large bulky hydrophobic residues at the P4 position. X-ray structures of SARS-CoV-2 PLpro show that the P2 binding sub-pocket is very narrow and can accommodate only a very small side chain such as H atom of glycine. Whereas the P4 binding pocket is wide and surrounded by several hydrophobic residues suitable for bulky hydrophobic side chains. Such findings led to the development of peptide inhibitors consisting of unnatural amino acids at the P3 and P4 positions and vinyl ester at the N-terminus as a warhead for the formation of covalent bond. These inhibitors covalently link to Cys-111 through thiether bond and perfectly fit in the substrate binding pocket as has been depicted by their X-ray crystal structures in complex with SARS-CoV-2 PLpro (Rut et al., 2020). These inhibitors, VIR250 and VIR251 (Fig. 2c), inhibit the catalytic activity of the enzyme and binding of ubiquitin and interferon stimulated gene 15 (ISG15) to SARS-CoV-2 PLpro and do not bind to any other protease, thereby showing high specificity for SARS-CoV-2 PLpro (Rut et al., 2020). Structural description of the binding interactions of these inhibitors provides essential information for the design of new pharmacophores and potent inhibitors of this enzyme. Recently, an FDA approved drug against leukemia, 6-thioguanine was reported to inhibit SARS-CoV-2 PLpro with IC50 of 0.1 to 0.5 μM and viral replication with EC50 of 0.647 ± 0.374 μM. However, the structural basis of this inhibition by 6-thioguanine remains to be characterized (Swaim et al., 2020).
In addition to processing the pp1a to produce functional nsp1, nsp2 and nsp3, SARS-CoV-2 PLpro can bind and hydrolyze ubiquitin and ubiquitin-like protein ISG15. These two cellular proteins contain the Leu-X-Gly-Gly cleavage motif of PLpro at their C-terminus. Ubiquitin and ISG15 are cellular regulatory proteins that can be post-translationally incorporated to a target protein through the formation of isopeptide bond between the C-terminus of these proteins and ε-amino group of lysine side chains of the target protein. PLpro can hydrolyze the isopeptide bonds removing ubiquitin and ISG15 from the host proteins therefore the enzyme is also referred to as isopeptidase. ISG15 and ubiquitin are signaling elements that play an important role in the innate immune response against viruses. Deubiquitinating and deISGylating activities of SARS-CoV-2 PLpro thus antagonize ubiquitin-dependent cellular response against the virus. In this regard, PLpro has been demonstrated to inhibit the production of cytokines involved in the activation of host innate immune response. It has been shown to interfere with the activation of transcriptional factor interferon regulatory factor-3 (IRF3) and NFκB signaling pathway thereby mediating an antagonistic effect against antiviral host immune response.

Interestingly, the catalytic activity of PLpro leading to deubiquitination and deISGylation of the host protein or just it’s binding to cellular protein-bound ubiquitin and ISG15 has been shown to mediate the antagonistic effect of viral PLpro. Therefore, inhibiting the catalytic activity of PLpro as well as interfering it’s binding to cellular protein-bound ubiquitin and ISG15 can restore cellular antiviral innate immune response against the virus. Inhibition of coronavirus PLpro thus, can lead to inhibit the viral replication and overcome the viral antagonistic effect on the host innate immune response against the virus. A small organic molecule GRL-0617 (snyder457) and its three derivatives (Fig. 2c) have been reported to bind to PLpro at the entrance of its catalytic groove and block the binding of its substrate. GRL0617 inhibits the viral replication and promotes the antiviral immunity (Shin and Mukherjee, 2020). These compounds fits into a binding pocket formed by the movement of Tyr-268 containing loop (Fig. 2b) and the binding is mediated primarily by hydrophobic interactions of naphthyl ring of these inhibitors and hydrophobic pocket formed by Tyr-268 and the two conserved proline residues, Pro-247 and Pro-248. In addition to the main catalytic pocket the pocket formed between Tyr-268 and Pro-247/248 represents an important potential hotspot for therapeutic intervention.

Structurally, PLpro exhibits multiple potential therapeutic hotspots. (A) X-ray crystal structure of SARS-CoV-2 PLpro shown in cartoon, with the illustration of fingers, palm, thumb and ubiquitin-like (UL) domains (PDB ID 6wuu). The catalytic residues are displayed with green sticks and the two ubiquitin-binding sites (UBS1 and UBS2) are colored blue and purple, respectively. (B) The structure of SARS-CoV-2 PLpro overlaid by its complex with the dual action inhibitor, GRL0617 showing the interactions of the inhibitor with Tyr-268 that has adopted different conformation in the complex structure (PDBID 7jir). (C) Chemical structures of different SARS-CoV-2 PLpro inhibitors.
96% sequence conservation. Two amino acid residues His-41 and Cys-145 involved in catalysis are highly conserved. In addition to these catalytic residues His-163 and Phe-140 facilitate the binding of glutamine at the P1 position of the peptide substrate through two H bonds between NZ2 atom of His-163 and backbone carbonyl oxygen of Phe-140 with NZ1 and NZ2 atoms of substrate glutamine, respectively (Xue et al., 2008). A typical oxyanion whole is formed between Cys-145 and Gly-143 to derive the catalysis (Fig. 3a). The substrate specificity of 3CLpro is primarily defined by the residues at the P1, P1′ and P2 positions of the peptide substrate. These positions are highly conserved in all coronaviruses in particular the presence of glutamine at the P1 position (N-terminus of the scissile bond) of the substrate is strictly required for 3CLpro binding across all coronaviruses (Fig. 3c) (Ullrich and Nitsche, 2020).

Most of the SARS-CoV-2 3CLpro inhibitors reported to date are covalent inhibitors where different warheads have been used to form covalent linkage with the thiol side chain of the key catalytic residue, Cys-145. In all these designed inhibitors, the lactam ring has been used as a surrogate of the P1 glutamine residue. Another common feature in the inhibitors reported by three different research groups is the presence of hydrophobic groups at the P2 site. Zhang et al., have reported the design of α-ketoamide derivatives by incorporating the ketoamide group as a warhead for covalent linkage of the inhibitor. These inhibitors form covalent bond with the side chain of Cys-145, resulting from a nucleophilic attack by the thiol group of Cys-145 on the α-keto group of the inhibitor to form thiohemiketal. One of the potent inhibitors of this class (Fig. 3b) shows IC₅₀ of 0.67 µM against purified SARS-CoV-2 3CLpro and EC₅₀ of ~4 µM in viral replication assay. The structure of this inhibitor in complex with 3CLpro shows that the lactam ring at the P1 position fits into the S1 pocket, cyclopropyl methyl side chain at the P2 position occupies position in the S2 sub-pocket and tert-butyloxycarbonyl (Boc), a protecting group on the P3 residue stabilizes the binding through hydrophobic interactions with proline-168. Boc has also been associated with its role in mediating transport of the inhibitor across the cell membrane (Zhang and Lin, 2020).

Dai et al., have reported a group of covalent inhibitors incorporating the aldehyde group as a warhead for covalent linkage. These compounds (11a and 11b) (Fig. 3d) also contain the lactam group at the P1 site that fits into the S1 subsite. The cyclohexyl ring or the 3-fluorophenyl group at the P2 site inserts in the S2 cavity and is stabilized primarily through hydrophobic interactions. The indole group at the P3 position is primarily stabilized by hydrophobic interactions. These compounds inhibit SARS-CoV-2 infection with IC₅₀ of 0.04 to 0.05 µM and show promising safety profile and pharmacokinetic properties (Dai and Zhang, 2020).

Another covalent inhibitor, N3 was reported by Jin et al., (Jin et al., 2020). This inhibitor forms a covalent bond with Cys-145 presumably through Michael addition reaction between the Cys-145 thiol side chain and Cβ of the vinyl group of the inhibitor. This inhibitor also contains the lactam ring at the P1 position that inserts in the S1 sub-pocket. The side chain of leucine at the P2 site is inserted in the S2 sub-pocket while at the P3 and P4 positions hydrophobic residues are present. This compound inhibits the SARS-CoV-2 replication with EC₅₀ of 16.8 µM. The same
group also identified a few small organic molecules through in silico screening of drug libraries followed by in vitro validation. The most potent compound among these was ebselen (Fig. 3d) that inhibited SARS-CoV-2 3CLpro with IC\textsubscript{50} of 0.67 µM and viral replication with EC\textsubscript{50} of 4.7 µM. The mass spectrometry analysis showed that these compounds were also covalently linked to the Cys-145 of 3CL-pro (Jin et al., 2020).

The only non-covalent inhibitor of SARS-CoV-2 3CLpro reported to date with sub-micromolar IC\textsubscript{50} is baicalein (Fig. 3d), a flavonoid of plant origin. Baicalein has been reported to inhibit SARS-CoV-2 3CLpro with IC\textsubscript{50} of 0.94 ± 0.2 µM and SARS-CoV-2 infection in the Vero cells with EC\textsubscript{50} of 2.94 ± 1.19 µM. X-ray crystal structure of the complex of this compound and 3CLpro shows it’s binding at the catalytic site of the enzyme. The binding is stabilized by several H bonds between three hydroxyl groups of the inhibitor and the backbone atoms of key residues of catalytic pocket including Gly-143, Ser-144, Cys-145, and hydrophobic interactions between it’s phenyl ring and the S2 sub-pocket (Sa et al., 2020b). All these inhibitors and structural description of their interactions with the enzyme provide insight into rational design of more potent inhibitors of 3CLpro as potential candidates to treat SARS-CoV-2 infections.

The 3CLpro enzyme forms a homodimer and it is catalytically active only in the dimer form (Pillaiyar et al., 2016; Zhang and Lin, 2020). Therefore, from inhibitor design perspective, in addition to the main substrate binding site the interprotomer contact surface represents a potential target site to interrupt the dimer formation leading to inhibit the catalytic activity of the enzyme (Fig. 3a). However, no example of such inhibitor of 3CLpro exists to date.

2.3. RdRp of SARS-CoV-2

RdRp of positive sense RNA viruses plays a central role in the viral replication cycle. RdRp catalyzes the synthesis of negative-strand RNA (−RNA), new genomic RNA and sub-genomic messenger RNAs during viral replication (Kim and Chang, 2013; Shereen et al., 2020; te Velthuis et al., 2012). Moreover, the N-terminal domain of the coronavirus RdRp exhibits nucleotidyltransferase enzymatic activity that also plays an important role in viral replication (Lehmann et al., 2015). RdRp shares 98% sequence identity between SARS-CoV-2 and SARS-CoV. The RdRp domain is present in the NSP12 subunit of the viral genome and is translated as a part of the pp1a polyprotein (Fig. 1), and successively cleaved from the polyprotein by proteolytic activity of the viral 3CLpro enzyme. Subsequently, RdRp is incorporated into a membrane-associated complex for its functions. The viral proteins that interact with RdRp to form the complex include NSP7, NSP8 and NSP13 (Chen et al., 2020; Subissi et al., 2014). Recently, several cryo-EM structures of SARS-CoV-2 RdRp have been reported. These structures include complex of RdRp with its cofactors nsps7 and nsps8 (Gao et al., 2020), RdRp-nsp7-nsp8 complex containing bound RNA (Hillen and Kokic, 2020), an inhibitor remdesivir (Yin et al., 2020) and nsps13 (Chen et al., 2020). The structure of RdRp-nsp7-nsp8 complex without bound RNA indicates that binding of nsps7 and nsps8 to RdRp is not primarily dependent on the bound RNA.

RdRp consists of 920 amino acid residues. Its structure adopts a typical canonical right-hand architecture containing the palm, the thumb and the fingers sub-domains, an interface domain, an N-terminus nidoivirus RdRp-associated nucleotidyl tranferase (NiRAN) and a β hairpin. Two highly conserved amino acid residues Asp-760 and Asp-761 are the main catalytic residues that are present in the palm sub-domain of the protein and the catalytic groove is surrounded by the palm, the thumb and the fingers sub-domains. In the RNA bound holoo-RdRp structure, two nsps bind to RdRp opposite to the catalytic groove and their helical extensions primarily consisting of positively charged residues protrude in parallel to the exiting RNA while binding to it (Hillen and Kokic, 2020). Nsp13-bound RdRp replication machinery contains two nsp13 molecules. The zinc-binding N-terminal domain of each of the nsp13 molecules interacts with the protruding helical extension of each copy of nsp8 while the ATPase domain of each helicase is positioned in front of holo-RdRp (Chen et al., 2020). The NiRAN domain contains Mg\textsuperscript{2+} and nucleotide-binding sites. The NiRAN domain might be involved in the activation of nucleoside triphosphate to be incorporated into the growing chain of RNA where Mg\textsuperscript{2+} plays a crucial role in the transferase activity of the domain. However, such role of the NiRAN domain remains to be experimentally confirmed.

Nucleotide analogues can mimic natural nucleotide substrates of the enzyme to be incorporated into the new RNA chain and inhibit further elongation of the RNA. Generally, three types of nucleotide analogue RdRp inhibitors have been reported in positive sense RNA viruses. The analogues that lack 3′–OH group, halt the RNA strand elongation after being incorporated (Mitsuuyai et al., 1990). Some nucleotide analogues do not halt the RNA strand synthesis but incorporate a permanent mutation in the RNA due to modified base, such compounds are called mutagens (Crotty et al., 2000). Some analogues contain natural base and 3′–OH of the ribose but the ribose ring is modified. These inhibitors after incorporation interrupt translocation as the termination of the RNA strand occurs after the incorporation of three more bases; such termination is referred to as non-obligate chain termination. Every type of nucleotide analogue inhibitors of RdRp needs to be in the form of nucleoside triphosphate to mimic the natural nucleotide substrates of the enzyme (Warren et al., 2014). Nucleoside triphosphates are less stable and have limited permeability across the cell membrane. Therefore, such drugs are developed in the form of pro-drugs that contain hydrophobic groups capping the polar phosphate to facilitate passage across the cell membrane. Inside the cell the hydrophobic groups are removed and subsequently triphosphate form representing the active drug is produced. Remdesivir that has been approved for a widespread phase-3 clinical trial to treat COVID-19 is an adenosine analogue that is incorporated into the RNA strand by RdRp resulting in non-obligate chain termination during RNA synthesis (Jorgensen et al., 2020; Sisay, 2020). The triphosphate active form of remdesivir has been shown to be more reactive to RdRp than its natural substrate ATP (Yin et al., 2020). Ribavirin, a previously approved antiviral drug, is also a nucleotide analogue RdRp inhibitor, and is under clinical trial in combination therapy against COVID-19 (Hung et al., 2020). Favipiravir, a broad-spectrum antiviral drug is a structural analogue of the guanine base and is the inhibitor of viral RdRp. This drug has also shown limited efficacy against SARS-CoV-2 in clinical use (Jomah et al., 2020). Sofosbuvir another nucleotide analogues has been demonstrated in silico to inhibit SARS-CoV-2 RNA replication (Jacome et al., 2020).

Multiple nucleotide analogue RdRp inhibitor drugs have shown efficacy against SARS-CoV-2 in clinical trials to varying extent and are at the early phases of development (Gadebusch Bondio and Marloth, 2020), however, no non-nucleoside SARS-CoV-2 RdRp inhibitor has been reported to date. Recent structures of RdRp in complex with RNA and different cofactor proteins provide essential insight into the rational design of RdRp inhibitors. In SARS-CoV-2 RdRp, besides the main catalytic groove containing catalytic sub-site around the Asp-760 and Asp-761 residues, there is a well defined ADP and Mg\textsuperscript{2+} binding site in the NiRAN domain. Although the precise role of the NiRAN domain in RNA transcription and replication remains to be elucidated yet this site represents a potential hotspot to be targeted for therapeutic intervention (Fig. 4). RdRp requires it’s binding to the nsps7 and nsps8 cofactor proteins for RNA replication process. SARS-CoV-2 RdRp exhibits two distinct nsps8 and one nsps7 binding sites (Fig. 4). These sites can be explored to identify pockets for inhibitors that could interrupt the binding of RdRp to the essential cofactor proteins. Similarly, complementary sites on nsps8 and nsps7 could also be exploited to interrupt their binding to RdRp as a potential strategy to discover new inhibitors of SARS-CoV-2 replication.
2.4. Helicase of SARS-CoV-2

Coronavirus helicase (NSP13) catalyzes the unwinding of duplex RNA into single strand nucleic acid chains during RNA transcription and replication (Snijder et al., 2016), the process is indispensable for viral replication and propagation. The process of nucleic acid unwinding is driven by the energy from hydrolysis of NTP, which is also catalyzed by helicase. SARS-CoV-2 helicase contains distinct RNA and NTP binding sites. In addition to these two important binding sites there is a \( \beta_{19}-\beta_{20} \) loop encompassing residues 331–357 that plays a crucial role during the nucleic acid unwinding process (Jia et al., 2019). However, a detailed mechanism of how the \( \beta_{19}-\beta_{20} \) loop plays a role in RNA duplex unwinding yet to be delineated. Helicase exhibits high sequence conservation among all coronaviruses (Jia et al., 2019; Mirza and Froeyen, 2020), in particular it shows 99.8% sequence identity between SARS-CoV-2 and SARS-CoV (Iftikhar et al., 2020). A typical CoV helicase consists of a zinc-binding domain, which is bisected by domains 1B and stalk from the main catalytic domains 1A and 2A.

Recently, cryo-EM structures of SARS-CoV-2 helicase and its complex with RNA-bound RdRp-nsp7-nsp8 transcription/replication machinery (holo RdRp) have been reported. In these structures, two nsp13 have been shown to bind to holo RdRp. In the structure, the N-terminal zinc-binding domain (ZBD) of each nsp13 interacts with the N-terminal helical protrusion of each copy of nsp8. ZBD and the catalytic domain 1A of one of the nsp13 molecules also make contacts with the thumb domain of RdRp and nsp8 head region, respectively. The amino acid residues involved in all these interactions are universally conserved among coronaviruses (Chen et al., 2020). Binding of nsp13 to RdRp complex has been shown to significantly enhance the NTPase and unwinding activity of the helicase enzyme in SARS-CoV (Jia et al., 2019). SARS-CoV-2 helicase structure incorporates five different potential hotspots that can be explored to interfere with its regular function and for therapeutic intervention. These hotspots include, distinct NTP and RNA binding sites, the \( \beta_{19}-\beta_{20} \) loop, the nsp8/RdRp binding site on ZBD and a site on the 1A catalytic domain that mediates interaction of nsp13 with the head region of nsp8 (Fig. 5).

In addition to the essential role in RNA transcription and replication, SARS-CoV-2 nsp13 has also been demonstrated to potently antagonize the interferon pathway and suppress the production of primary interferon. Interferon plays an essential role in the host defense against viruses. Nsp13 inhibits the nuclear localization of interferon regulatory factor-3 (IRF3) that acts as transcriptional factor of interferon \( \beta \) (Yuen et al., 2020). Nsp13 therefore, inhibits the production of primary interferon and interferon pathways. However, structural description of nsp13 interactions with any of the factors of interferon pathway yet to be described. Intervening the interferon antagonizing effect of nsp13 could potentially restore the antiviral immune response and suppress the viral propagation.

Although the coronavirus helicase enzyme incorporates several
potential therapeutic hotspots but only a few inhibitors of this enzyme have been reported. Shum et al have reported the synthesis of different DNA aptamers, which inhibit DNA unwinding activity of SARS-CoV helicase in vitro with IC\textsubscript{50} ranging from 87 to 120 nM (Shum and Tanner, 2008). Purine derivatives have been shown to inhibit ATPase as well as RNA-duplex unwinding activity of SARS-CoV helicase with IC\textsubscript{50} of 8.6 and 41.6 \mu M (Cho et al., 2015). Mirza et al., recently performed extensive in silico screening followed by molecular dynamics simulation to identify three aromatic compounds as novel potential inhibitors of SARS-CoV-2 helicase (Mirza and Froeyen, 2020). White et al., recently identified two previously known drugs lumacaftor and cepharamine to inhibit SARS-CoV-2 Nsp13 ATPase activity with in vitro IC\textsubscript{50} values of 0.3 and 0.4 mM (White et al., 2020). The structural description of all these inhibitors and their interactions with SARS-CoV-2 helicase will provide basis for the rational design of new potent inhibitors of helicase as potential antiviral agents.

3. Host factors

Targeting host factors involved in key steps of viral pathogenesis has advantage of less chances of the emergence of drug-resistant viral variants, particularly in genetically variable viruses (Edinger et al., 2014). However, on the contrary, targeting a host factor may result in significant modulation in the associated physiological function leading to side effects. Nevertheless, to contain a life threatening viral infection by targeting a host factor the associated side effects could be compromised.

3.1. ACE2 a potential target for therapeutic intervention against SARS-CoV-2

ACE2 is a key component of the renin angiotensin system (RAS) that contributes to maintain the blood pressure and balance of electrolytes thereby playing an important role in cardiovascular and kidney functions. Renin cleaves angiotensinogen to produce angiotensin I, which is further cleaved by angiotensin converting enzyme (ACE) into angiotensin II. Angiotensin I and II are further processed by ACE2 to produce angiotensin 1–9 and angiotensin 1–7, respectively. Angiotensin 1–7 stimulates vasodilation while angiotensin II acts as a vasoconstrictor. Therefore, a balance between angiotensin II and angiotensin 1–7 is important to maintain the homeostasis (Imai et al., 2010; Kuba et al., 2013). ACE2 is a transmembrane glycoprotein with 120 kDa molecular mass with an extracellular catalytic domain. ACE2 is expressed in several types of tissue including lungs, heart, kidneys, testes and colon.

Enzymatically, ACE2 acts as a carboxypeptidase and exhibits the conserved active site HEXXH motif for the enzymatic activity. ACE2 was identified as the primary receptor of SARS-CoV in 2003 (Li et al., 2003) and further studies suggested that binding of viral spike protein to ACE2 did not block its catalytic activity (Kuba et al., 2013). SARS-CoV-2 binds to ACE2 with roughly 10 fold higher affinity as compared to SARS-CoV explaining the basis of much higher transmissibility of SARS-CoV-2 (Tai et al., 2020). Recently, the crystal structure of the complex of ACE2 and the receptor-binding domain (RBD) of SARS-CoV-2 spike has been solved by Shang et al., that delineates the key interactions mediating the viral recognition of its ACE2 receptor (Shang et al., 2020). Interference with these interactions by targeting either of the interacting proteins could inhibit the viral entry to the target cell. Viral spike RBD-binding site on ACE2 is rather shallow and wide without a very well defined pocket (Fig. 6a). However, small pockets within the wide binding site could be identified and targeted to interfere with the interactions between the viral spike and ACE2.

Recent molecular dynamics study has shown that peptides surrogate of the spike-binding motif of ACE2 inhibits the interaction of the S protein to ACE2 by binding to the S protein (Baig et al., 2020). Monteil et al., have shown that a clinical grade recombinant soluble ACE2
functions. Moreover, several inhibitors of ACE2 carboxypeptidase activity are known that inhibit ACE2 catalyzed proteolysis (Takahashi et al., 2015). Such inhibitors are less likely to intervene the binding of SARS-CoV-2 to the ACE2 receptor as the ACE2 catalytic site is in a distinct structural region than the S-protein binding site (Fig. 6a).

3.2. TMPRSS2 as a potential target for therapeutic intervention against SARS-CoV-2

SARS-CoV-2 cellular entry is mediated by the binding of its S protein to the ACE2 receptor and subsequent priming of the S protein by cellular protease that involves the cleavage of the S protein at two different sites namely S1/S2 and S2′. Such priming enables the fusion of viral membrane to the host cellular membrane. Transmembrane serine protease 2 (TMPRSS2) catalyzes this proteolytic cleavage of the S protein (Hoffmann et al., 2020). In the S1/S2 cleavage site of SARS-CoV-2 there is an insertion of four residues (PRRA) as compared to SARS-CoV – the S1/S2 cleavage motifs of SARS-CoV and SARS-CoV-2 consist of SLIR|S and RRAR|S, respectively while the S2′ cleavage site in the two viruses is identical. SARS-CoV requires cysteine protease, cathepsin for the cleavage of S1/S2 site and TMPRSS2 for the cleavage of S2′ site, while in SARS-CoV-2 cleavage at both sites occurs by TMPRSS2 (Hoffmann et al., 2020). This difference has been suggested as the basis of higher transmissibility of SARS-CoV-2 as compared to SARS-CoV (Wang et al., 2020c). TMPRSS2 is a member of transmembrane serine proteases. Different members of this family are expressed in a variety of tissues in a tissue dependent manner. TMPRSS2 has also been reported to play an important role in the cellular entry of avian influenza viruses through priming of hemagglutinin of these viruses (Shin and Seong, 2017). Therefore, TMPRSS2 could be targeted for the development of broad-spectrum antiviral agents.

TMPRSS2 consists of an N-terminal cytoplasmic domain, a hydrophobic transmembrane domain and an extracellular scavenger receptor domain followed by the C-terminal serine protease domain. The serine protease domain contains the Ser-His-Asp catalytic triad where the serine residue acts as a nucleophile during catalysis. Recently, Hoffmann et al., described the inhibition of SARS-CoV-2 and SARS-CoV in lungs cell line Calu-3 and in primary human lungs cell by camostat mesylate, a synthetic serine protease inhibitor previously approved for the treatment of chronic pancreatitis and oral squamous cell carcinoma (Hoffmann et al., 2020). Similarly another serine protease inhibitor nafamostat mesilate (Fig. 6d) which was approved in Japan to treat acute pancreatitis, has been shown to inhibit SARS-CoV-2 entry to the Vero cells with EC50 of 22.5 μM (Wang et al., 2020b). Although these drugs have shown viral entry inhibition in vitro, their efficacy to treat COVID-19 remains to be studied. As the structure of TMPRSS2 is not available yet, to gain an understanding of its structural features we homology modeled the structure of extracellular domain of TMPRSS2 using serine protease hepsin (ID: 5ce1.1.A) as template that showed 34% sequence identity. The structure of extracellular portion incorporates a scavenger receptor domain and a typical serine protease domain. The putative substrate-binding site with catalytic triad has been depicted in Fig. 6c. The TMPRSS2 modeled structure could be used to rationally design its inhibitors until its cryo-EM or X-ray structure is available.

3.3. Endoplasmic reticulum (ER) glucosidas as potential targets for therapeutic intervention against SARS-CoV-2

The S protein of SARS-CoV-2 is highly glycosylated as each protramer of the S protein trimer incorporates 22 N-glycosylation sites. Watanabe et al., have recently reported the mass spectrometry based glycan profiling of the recombinant S protein of SARS-CoV-2. According to this profiling, SARS-CoV-2 exhibits different glycoforms including high-mannose, complex and hybrid type glycans covering the whole surface of the S protein (Watanabe et al., 2020). Senda et al., have reported the presence of 90-linked glycans on the recombinantly expressed S protein, these glycosylation sites also include Thr-678 that is present in the vicinity of TMPRSS2 cleavage site on the S protein and is unique to SARS-CoV-2 (Senda et al., 2020). Co- and post-translationally incorporated glycans on viral proteins play diverse roles in viral pathobiology. Glycans mediate protein folding and stability, and play important roles in defining viral tropism and immune evasion by shielding the protein epitopes from the immune system (Dalziel et al., 2014; Watanabe et al., 2019). In the SARS-CoV-2 S protein no mutation of any N-glycosylation site has been reported since the emergence of this infection suggesting essential roles of glycans in the SARS-CoV-2 pathogenesis and survival (Watanabe et al., 2020).

Intervening the glycosylation process of viral proteins also represents an important strategy for therapeutic intervention in particular for highly glycosylated viruses such as coronaviruses. N-glycosylation represents the most common type of glycosylation in eukaryotes that occurs in the lumen of endoplasmic reticulum (ER) by the incorporation of a preformed glycan precursor Glc3Man9GlcNAc2 to the asparagine residue present in the NXS/T sequon where “X” represents any residue except proline. The incorporated glycan is processed by a series of enzymes present in the ER and the Golgi apparatus. Initial steps of this processing involve stepwise trimming of the three glucose saccharides from one of the terminal branches of Man9-GlcNAc2 by the α-glucosidas I and α-glucosidas II enzymes, in succession (Fig. 7). The trimmed Man9-GlcNAc2 or Man8-GlcNAc2 is transported to the Golgi apparatus for further glycan processing.

Inhibition of ER α-glucosidas represents a potential approach to discover new antiviral agents. Celgosivir, the precursor of a known ER α-glucosidase I inhibitor castanospermine, has shown efficacy against hepatitis C virus (HCV) infections and this potential drug underwent phase II clinical trials in combination therapy with pegylated interferon-α and ribavirin to treat hepatitis C (Durantel, 2009) – HCV envelope protein E2 exhibits 11 N-glycosylation sites (Iacob et al., 2008). Chapel et al., has reported the effect of ER glucosidase inhibitors on HCV morphogenesis. HCV virus like particles (VLPs) produced in the presence of ER α-glucosidase inhibitors contained primarily unprocessed glycans and partially miss-folded envelope protein (Chapel et al., 2006). Ma et al., recently reported the inhibition of different hemorrhagic viruses including the yellow fever and ebola viruses by an ER α-glucosidase I inhibitor, IHVR-19029 in vitro and in animal models when used in combination with favipiravir (Ma et al., 2018). In 2014 two siblings were reported with a rare genetic disorder of glycosylation type Ib resulting from mutations in the gene encoding ER glucosidases. These siblings were resistant to viral infections indicating key roles of ER glucosidoses in viral morphology and pathogenesis (Sadat et al., 2014). Effect of different immunosugar inhibitors of different glycan processing enzymes of the ER and the Golgi apparatus on the infectivity of different hemorrhagic viruses in vitro and in mouse models have been described (Tyrell et al., 2017). ER α-glucosidase inhibitors not only can alter the glycosylation processing of the viral glycoproteins leading to affect viral fitness but also can alter the morphology of viral receptor proteins if those are glycosylated. Zhao et al., have shown that α-glucosidase I inhibitor, IHVR-17028 inhibits the entry of lentiviral particles pseudo-typed with envelopes of SARS-CoV and human coronavirus NL-63 due to the altered glycosylation of ACE2 which is a glycoprotein (Zhao et al., 2015).
Therefore, inhibition of ER glucosidases can potentially inhibit the SARS-CoV-2 infections through dual mechanisms; by altering the glycosylation of the viral S protein leading to impact its morphogenesis and by altering the glycosylation of the SARS-CoV-2 receptor ACE2 leading to impair its proper structure resulting in an effect on viral-receptor recognition and subsequent entry. Most of the known inhibitors with antiviral effects are the derivative of well-known glucosidase inhibitors 1-deoxynojirimycin (DNJ) and castanospermine (CAST) (Fig. 7). These derivatives have shown inhibition of different enveloped viruses. However, studies in animal models show rather moderate inhibition of viral infections by these drugs suggesting a need of designing more potent inhibitors. Chang et al., has provided a comprehensive review on different glucosidase inhibitors and their antiviral effects in detail (Chang et al., 2013).

The human ER glucosidases structures are not yet available, however, structures of glucosidase I from yeast and glucosidase II from different organisms have been reported (Barker and Rose, 2013; Caputo et al., 2016). Solving the structure of human ER glucosidases through cryo-EM or X-ray crystallography will provide detailed insight into the rational design of more potent and specific ER glucosidase inhibitors. Therefore, inhibition of ER glucosidases can potentially inhibit the SARS-CoV-2 infections through dual mechanisms; by altering the glycosylation of the viral S protein leading to impact its morphogenesis and by altering the glycosylation of the SARS-CoV-2 receptor ACE2 leading to impair its proper structure resulting in an effect on viral-receptor recognition and subsequent entry. Most of the known inhibitors with antiviral effects are the derivative of well-known glucosidase inhibitors 1-deoxynojirimycin (DNJ) and castanospermine (CAST) (Fig. 7). These derivatives have shown inhibition of different enveloped viruses. However, studies in animal models show rather moderate inhibition of viral infections by these drugs suggesting a need of designing more potent inhibitors. Chang et al., has provided a comprehensive review on different glucosidase inhibitors and their antiviral effects in detail (Chang et al., 2013).

Although ER α-glucosidase inhibition is a potential approach for antiviral drug discovery, however, such inhibition is expected to alter the glycosylation of every cellular glycoprotein leading to impart adverse physiological effects. Therefore, tissue specific cellular marker targeted drug delivery can minimize the off target effects of host targeted drugs, in general (Zhao et al., 2020). Recently, Bouhaddou et al., reported a mass spectrometry based phosphoproteomics and scanning electron microscopy based morphological analyses of the SARS-CoV-2 infected cells and showed an altered morphology of the infected cells with filopodial protrusions on the cellular surface (Bouhaddou et al., 2020). Such morphological exclusiveness of SARS-CoV-2 infected cells and associated cell surface markers can be exploited for an infected cell specific drug delivery. Warfield et al., very recently reported that single dose of a derivative of DNJ, N-(9′-methoxyynonyl)-1-deoxynojirimycin, an ER α-glucosidase inhibitor, prevented the death of mice infected with lethal doses of dengue and H1N1 influenza viruses. The single dose of the drug showed extended pharmacological effect in animal models and safety in human subjects at a dose as high as 1 g (Warfield et al., 2020).

Such findings highlight a novel paradigm for the development of treatments against acute viral infections.

4. Conclusions

SARS-CoV-2 infections have emerged as one of the most formidable health risks and are adversely affecting economies, healthcare, educational system, social and personal lives all over the world. The COVID-19 cases continue to increase in many parts of the world despite vigorous and collective global efforts of its containment. Since the emergence of this pandemic substantial multidisciplinary and cohesive research efforts have resulted in the creation of unprecedented knowledge about the virus and its cross talk with the host. Molecular structures of various SARS-CoV-2 proteins and their complexes with interacting partners explaining the structural basis of viral pathogenesis and its replication cycle have become available as a result of striking progress in technology and extraordinary efforts of the scientific community. In this review we recapitulated the structural information of four viral factors, RdRp, 3CLpro, PLpro and helicase and two host factors including ACE2 and TMPRSS2 in the context of explaining potential hotspots to target for drug development against SARS-CoV-2 in addition to discussing ER glucosidases as potential drug targets. Future efforts should be made to identify new potential inhibitors of the SARS-CoV-2 infections by exploring different potential target sites on proteins in addition to their main catalytic sites. Moreover, solving the structures of human ER glucosidases would also be an important future perspective for rational drug design. In brief, by combining the emerging structural information, computer-aided drug design and medicinal chemistry techniques, rational drug design efforts should result in the development of a number of potential drugs against SARS-CoV-2.

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.

Acknowledgements

We highly acknowledge the Department of Biology, SBA School of Science and Engineering at Lahore University of Management Sciences for a continues support of research in the form of resources.
Authors contributions
MS wrote a major part of the original draft, SS provided assistance in manuscript preparation, formal analysis, and reviewing & editing.

References
Al-Kasasbeh, J., Pederson, J., Kohberger, G., 2020. Vaccine Candidates against Coronavirus Infections. Where Does COVID-19 Stand? Viruses 12, 861–879.
Baig, M.S., Alagamuthu, M., Rajpoot, S., Saqib, U., 2020. Identification of a Potential Peptide Inhibitor of SARS-CoV-2 Targeting its Entry into the Host Cells. Drugs in R&amp;D 20, 161-169.
Barker, M.K., Rose, D.R., 2013. Specificity of Processing α-galactosidase I is guided by the substrate conformation: crystallographic and in silico studies. J. Biol. Chem. 288, 12563-12574.
Bayhan, J., Fenichel, E.P., 2020. Impact of school closures for COVID-19 on the US health-care workforce and net mortality: a modelling study. The Lancet. Public health 5, e271–e278.
Bouhaddou, M., Memon, D., Meyer, B., White, K.M., Rezelj, V.V., Correa Marrero, M., 2014. Emerging approaches to combat antiviral drug resistance. J. Gen. Virol. 95, 263–273.
Burkle, F.M., 2020. Declining Public Health Protections within Autocratic Regimes: Where Does COVID-19 Stand? Viruses 12, 861-879.
Chapel, C., Garcia, C., Roingeard, P., Zitzmann, N., Dubuisson, J., Dwek, R.A., Trepo, C., Chang, J., Block, T.M., Guo, J.-T., 2013. Antiviral therapies targeting host ER α-glucosidase II capture the binding modes of broad-spectrum α-glucosidase inhibitors on viral membrane proteins. Antiviral Res 108, 343-356.
Chen, J., Malone, B., Llewellyn, E., Grasso, M., Shelton, P.M.M., Olinares, P.D.B., Edinger, T.O., Pohl, M.O., Stertz, S., 2014. Entry of influenza A virus: host factors and origins and receptor binding. Lancet (London, England) 395, 1695-1704.
Edinger, T.O., Pohl, M.O., Stertz, S., 2014. Entry of influenza A virus: host factors and origins and receptor binding. Lancet (London, England) 395, 1695-1704.
