Chapter 2

Chemistry and Biology of SARS-CoV-2

This chapter is published
Alexander Dömling * and Li Gao^  
*Chem, 2020, 6(6), 1283-1295.
^Second authorship
Abstract

SARS-CoV-2 caused an unprecedented fast spreading worldwide pandemic. Also, currently with rather low mortality rate the virus spreads fast over the world by using traffic highways of our modern world. The coronavirus family members were responsible for several deadly outbreaks and epidemics in the last decade. Not only governments but also the scientific community reacted promptly to the outbreak and information is shared rather fast. For example, the genetic fingerprint was shared and the 3D structure of key proteins was rapidly solved which can be used for the discovery of potential treatments. An overview is given on the current knowledge of the spread, disease course, and molecular biology of SARS-CoV-2. We discuss potential treatment developments in the context of recent outbreaks, drug repurposing and development timelines.

SARS-CoV-2 • drug repurposing • protease • target • antiviral.
1. SARS-CoV-2 Outbreak in Wuhan

In December 2019 an outbreak of pneumonia of unknown cause was reported in Wuhan, in Hubei province, China. It was speculated that the first patients caught the infections from a seafood market that also traded wild animals. The causing agent was rapidly identified as a novel coronavirus. The coronavirus responsible for the outbreak is now called SARS-CoV-2. The respiratory illness caused by SARS-CoV-2 is called COVID-19. The symptoms of the SARS-CoV-2 infection range from asymptomatic to mild to severe to death. It soon became clear that person-to-person transmission was also occurring. In an unprecedented documented speed, the SARS-CoV-2 travels around the globe and led as of April 20th to >2.0 million infections and 150,000 fatalities. Dramatic measures were taken by the Chinese government and several multimillion inhabitant cities were isolated and put under quarantine in order to slow the pandemic spread. Different hosts of SARS-CoV-2 are proposed including snails, bats and pangolins.

Coronaviruses are a large family of zoonotic viruses, and common to humans. Under the electron microscope they exhibit pictures which are reminiscent of the solar corona. For example, the common cold is often caused by human coronaviruses. They are single-stranded enveloped RNA viruses and stand out through their rather large genome. As with viruses in general the structure is rather simple. SARS-CoV-2 is generally less pathogenic than SARS-CoV, and much less than MERS-CoV. The reported case-fatality rate of COVID-19 is ≤3% and is thus rather low as compared to SARS (30%, Table 1). However, the transmission rate (TR: number of newly infected people per infected person) is high with 2.5 to 3 and accounts for the danger of the current pandemic. For comparison the TR of the yearly common cold is less than 1.4.

Advice guidelines for diagnosis and treatment of SARS-CoV-2 infected pneumonia have been shared rapidly.

Table 1. Some data to the new and previous epidemi- and pandemics.

<table>
<thead>
<tr>
<th>pandemic (causing agent)</th>
<th>Transmission</th>
<th>Mortality</th>
<th>No. of infections</th>
<th>No. of deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black Death (Yersinia pestis) 14th century</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>&gt;50 million</td>
</tr>
<tr>
<td>Spanish flu (Influenza virus) 1918-1920</td>
<td>NA</td>
<td>10-20%</td>
<td>500 million</td>
<td>40-100 million</td>
</tr>
<tr>
<td>SARS (SARS-CoV) 2002-2003</td>
<td>NA</td>
<td>30%</td>
<td>8000</td>
<td>800</td>
</tr>
<tr>
<td>Covid-19 (SARS-CoV-2) Since Nov. 2019</td>
<td>&gt;2%</td>
<td>&lt;3%</td>
<td>&gt;2,000,000</td>
<td>&gt;150,000</td>
</tr>
<tr>
<td>Swine flu (H1N1 virus) 2009-2010</td>
<td>NA</td>
<td>0.01-0.1%</td>
<td>700-1400 million</td>
<td>50,000-575,000</td>
</tr>
</tbody>
</table>

What are the issues and chances for a rapid approval of a new medicine to treat COVID-19? In principle there are several potential strategies to pharmacologically fight COVID-19: vaccines, monoclonal antibodies, oligonucleotide-based therapies, peptides, interferon therapies, small-molecule drugs or natural medicines (e.g. traditional Chinese medicine, TCM). The timelines for the de novo development of a small molecule drug are historically > 6-7 years and in the best case less than 2 years. Vaccines can be developed much faster but still development in the range of 1-2 years is very challenging. Antibodies to support the body’s immune system are also a strategy to combat viral diseases. Again, the typically development timelines are several years. So, no hope for a rapid drug to come to the market? A strategy which is promising in the current situation is drug repurposing. Drug repurposing
aims to discovery novel indication areas for already approved drugs.\textsuperscript{4} The overwhelming advantage of drug repurposing is the potential much faster approval due to already extensive knowledge of the behavior of the drug in humans.

An expert opinion on the potential for repurposing existing antiviral agents to treat COVID-19, some of which are already clinically evaluated was recently given.\textsuperscript{5} Here we discuss molecular targets of the SARS-CoV-2, some of the known small molecules and the potential for repurposing existing drugs.

2. Molecular Biology and Targets

Despite the rather large size of the RNA virus genome of approx. 30,000 bases, the SARS-CoV-2 genome presumably encodes for few proteins (Fig. 1). the structural spike (S) protein, nucleocapsid (N) protein, membrane (M) protein, and the envelope (E) protein are required to produce a structurally complete viral particle. Additionally, the SARS-CoV-2 genome encodes non-structural proteins such as 3-chymotrypsin-like protease, papain-like protease, helicase, and RNA-dependent RNA polymerase.

![Figure 1. Schema of SARS-CoV-2 and its molecular protein targets.](image)

2.1. 3C-like protease (3CLpro)

Both the virus encoded proteases are involved in the processing of the viral polyprotein in a coordinated manner and thus comprise important drug targets. The structure and function and inhibition of coronavirus 3C-like protease (also called Mpro) has been recently comprehensively reviewed.\textsuperscript{6} The 3CLpro is a cysteine protease which cleaves and processes the viral polyproteins.
Figure 2. 3D-structure of SARS-CoV-2 3C-like protease bound to a covalent peptidomimetic inhibitor (PDB ID 6LU7). The active site Cys145 is indicated as yellow surface.

SARS-CoV-2 and SARS-CoV share 96% sequence identity in their 3CLpro. Based on the virus sequence data published rapidly a homology model was created. Moreover, an X-ray structure of the C3Lpro covalently bound to a peptidomimetic acrylester 1 is now available (Fig. 2, PDB ID 6LU7).

Due to the high sequence similarity of different corona virus 3C-like proteases a lot of previously described inhibitors can be considered of great use in the current SARS-CoV-2. The majority of inhibitors of the 3C-like protease are covalent in nature binding to the active site cysteine (Scheme 1).

Different electrophilic warheads are known including α-halocarbonyl, acrylamides, sulfonyl chlorides, aldehydes (2), isatines (3) or α-ketoheteraromates (4). Many of the molecules are rather large and based on extensive amide chemistry, mimicking part of the peptide substrate of the protease. Moreover, their selectivity towards other potential targets in the human body has not been established.
Scheme 1. Selected classes of 3C-like protease inhibitors. Warheads interacting covalently with the active site Cys145 are indicated in red.

Interestingly also some compounds binding to the active site of the 3C-like protease, however using a non-covalent mechanism have been established. A high throughput screening (HTS) identified 3,3-dihydropyrazolidine (5) which displayed 1,3,5-triaryl substitution patterns, as SARS-CoV 3CLpro inhibitors.\textsuperscript{11} Nitroanilides (6) derived from the drug niclosamide were found to inhibit 3CLpro.\textsuperscript{12} α-Aminoacylamides were found by a HTS and a strong stereochemical effect was noted. The simple one-pot accessible scaffold by an Ugi-four component condensation was key to rapidly generate SAR for the putative P2–P1 and P1 subgroups. An optimized version ML188 (7) was designated as the probe status (Fig.3). A P3 truncated version of 8 allowing for significant molecular weight (MW) reduction without diminishing potency was developed as a second probe ML300 (9) with potent enzyme inhibition and cellular activity. These compounds comprise rare examples of a noncovalent SARS-CoV 3CLpro inhibitor of moderate MW with good enzyme and antiviral inhibitory activity. However, these molecules suffer from extensive metabolism and rapid clearance. Nonetheless, they are a promising starting point for further drug development.
Figure 3. Non covalent probes binding to the active site of the 3C-like protease. Above: Cocrystal structure of probe ML188 bound to SARS-CoV (PDB ID 3V3M). The homology and the 3D structural similarity of SARS-CoV (green cartoon) and SARS-CoV-2 (pink cartoon) 3CL protease are very high. Close up view into the ML188 (cyan sticks) binding site. Below: Synthesis and some molecular, DMPK and pharmacology data of the probes ML188 and ML300.

Even if such compounds cannot be rapidly developed to cope with the current situation, their development is highly warranted to be prepared for likely future coronavirus outbreaks. Noteworthy, different computational approaches including machine learning were published to propose approved drugs potentially binding to 3CLpro (drug repurposing).13,14 One such approach virtually screened commercial medicines in the DrugBank database for binding into the active site of Mpro.15 Ten different commercial medicines were proposed that may form hydrogen bonds to key residues within the binding pocket of SARS-CoV-2 3CLpro, which may have higher mutation tolerance than lopinavir or ritonavir.
Flavonoids (10-14) are plant derived natural products with diverse reported biological activities and which have been shown to be also able to inhibit the 3CLpro (Fig.4).\textsuperscript{16,17} The broad spectrum and established use of plant-based medicines to combat infectious diseases in TCM is the basis of several currently ongoing clinical trials in China. One of the largest among them assesses shuanghuanglian, a Chinese herbal medicine that contains extracts from the dried fruit lianqiao (Forsythiae fructus), which is purported to have been used for treating infections for more than 2,000 years.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{flavonoids.png}
\caption{Flavonoids inhibiting 3C-like protease. Below: Liang Quiao, the seeds of the Forsythiae fructus plant used in the traditional Chinese medicine shuanghuanglian.}
\end{figure}

Several approved HIV protease inhibitors (15, 16, 18) were previously repurposed for the treatment of SARS (Scheme 2).\textsuperscript{18-20} They were hypothesized to inhibit the SARS-CoV 3CLpro: HIV protease is a Asp protease and differs from the Cys protease 3CLpro, but it also shares some common elements, like a tetrahedral transitions state and the receptor pockets to recognize the amino acid side chains of the substrates. Since SARS-CoV-2 and SARS-CoV share very high sequence identity in their 3CLpro, these HIV protease inhibitors are currently again repurposed for the treatment of COVID-19. (Chinese Clinical Trial Registry: ChiCTR2000029539)\textsuperscript{21-23}
2.2. Papain-like protease (PLpro)

The coronaviral papain-like protease (PLpro) is another attractive antiviral drug target because it is essential for coronaviral replication. The structure, function and inhibition of the SARS-coronavirus PLpro has been extensively reviewed.\textsuperscript{24} Although the primary function of PLpro and 3CLpro are to process the viral polyprotein in a coordinated manner, PLpro has the additional function of stripping ubiquitin and ISG15 from host-cell proteins to aid coronaviruses in their evasion of the host innate immune responses. Therefore, it was recently argued that targeting PLpro with antiviral drugs may have an advantage in not only inhibiting viral replication but also inhibiting the dysregulation of signaling cascades in infected cells that may lead to cell death in surrounding, uninfected cells.\textsuperscript{24}

Different compounds forming a covalent bond to the active site Cys112 have been described including epoxyketones, α-halo-ketones, alkynes, aldehydes, trifluoromethyl ketones, α, β-unsaturated ketone, activated esters or vinyl sulfones.

Disulfiram (19), an approved drug for the treatment of chronic alcohol dependence has a great potential for drug repurposing since it has been shown to inhibit PLpro of MERS and SARS-CoV.\textsuperscript{25} The antimetabolites 6-mercaptopurine (20) and 6-thioguanine (21) are additional drugs inhibiting the PLpro.
Figure 5. Cocrystal structure of HTS hit 22 (cyan sticks) with PLpro (grey sticks, PDB ID 3MJ5) showing some key interaction to Tyr269 forming a hydrogen bond and a T-shaped pi stacking interaction with the naphthyl moiety. 2D structures, biological activity and half life time of 22 and 23.

Potent naphthyl methylamine hits (22) were identified from a HTS campaign, structurally characterized and have been subsequently optimized towards better metabolic stability (23). The naphthyl methylamines work through a non-covalent mechanism and show a rather drug-like appearance (Fig.5).

2.3. Spike glycoprotein (S)

The envelope-anchored spike protein mediates coronavirus entry into host cells by first binding to a host receptor and then fusing viral and host membranes. The SARS-CoV-2 spike protein was solved by cryo-electron microscopy and just released. Knowing the Å-resolution structure of the SARS-CoV-2 spike will allow for additional protein engineering efforts that could improve antigenicity and protein expression for vaccine development. Moreover, the Å-resolution detail will enable the design and screening of small molecules with fusion-inhibiting potential.

It is the affinity between the viral receptor binding domain (RBD) and the host receptor in the initial viral attachment step that primarily determines which host is susceptible to SARS-CoV infection. The SARS-CoV-2 entry through receptor binding was elucidated independently by several groups. Based on the rich knowledge about SARS-CoV and the sequence homology, it was suggested that SARS-CoV-2 uses angiotensin-converting enzyme 2 (ACE2) as its receptor (Fig.6). It uses the SARS-coronavirus receptor angiotensin-converting enzyme 2 (ACE2) and the cellular protease TMPRSS2 for entry into target cells. The interplay of the ACE receptor in cardiovascular diseases (with the well known drug class of ACE inhibitors) and as the docking point for SARS-CoV-2 cellular infection is particular point of debate and research.

It was found that several critical residues in SARS-CoV-2 RBM (particularly Gln493) provide favorable interactions with human ACE2, consistent with 2019-nCoV’s capacity for human cell infection.
Figure 6. Structure of human SARS-CoV RBD complexed with human ACE2 (PDB ID 2AJF) as a model for SARS-CoV-2-ACE2 interaction. ACE2 is shown in blue, the RBD (receptor-binding domain) of SARS is in grey, and RBM (receptor-binding motif) of SARS-CoV in magenta. Closeup is highlighting specific hydrogen bindings of a Tyr cluster of SARS-CoV in cyan.

Based on the extensive modelling of the virus-human receptor interactions it was also predicted that a single N501T mutation may significantly enhance the binding affinity between SARS-CoV-2 RBD and human ACE2 and thus potentially lead to much more virulent bodies.\textsuperscript{30,31} The emergency and evolution of novel mutations at the S01 position in SARS-CoV-2 infected and Covid-19 patients has to be closely monitored.\textsuperscript{32}

On the other hand, the spike-shaped protein on the surface of the viruses causing SARS, MERS and COVID-19 provides a tantalizing target for antibodies or other compounds, which could prevent coronaviruses from invading human cells.\textsuperscript{35} The virus' spike protein seems to emerge as the consensus target antigen.

2.4. RNA-dependent RNA polymerase (RdRp)

The RdRp enzyme allows the viral genome to be transcribed into new RNA copies using the host cell's machinery. RdRp inhibitors are emerging as a new strategy to fight viral infections.\textsuperscript{35} The chemistry and biology of RdRp has been extensively reviewed.\textsuperscript{36}

RNA polymerase inhibitors are promising agents to fight Covid-19.\textsuperscript{3} RdRp proteins of SARS-CoV-2 and SARS-CoV share 96% sequence identity. Favipiravir (24) is an approved RNA polymerase inhibitor for the treatment of pandemic influenza. It has been shown not only to be active in influenza but also active against other RNA viruses (Fig.7).\textsuperscript{34} Favipiravir is a prodrug which is metabolized in cells into the active purine nucleotide mimic favipiravir-ribofuranosyl-5-triphosphate which inhibits the RNA replication and thus the viral progression.\textsuperscript{37} Interestingly, it does not inhibit host DNA and RNA synthesis and inosine 5'-monophosphate dehydrogenase (IMPDH) activity. Favipiravir reportedly
demonstrated efficacy with minor side effects in an ongoing 70-patient clinical trial in Shenzhen, Guangdong province. The drug’s generic version received the approval by health authorities in China.

The broad-spectrum antiviral drug Remdesivir together with chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. Remdesivir (25) reduced SARS-CoV-2 infection of monkey kidney cells with an EC50 of 0.77 μM. The compound is in late stage clinical development and has been recently described to inhibit multiple RNA viruses on a cellular level including Ebola and SARS. The presumed mode-of-action of the adenosine analog prodrug Remdesivir is pre-mature RNA synthesis termination by incorporation into nascent viral RNA chains. Galidesivir (26) is another antiviral drug under clinical development with potential in COVID-19 treatment (Fig.7). It is an adenosine analog and currently developed for Ebola virus disease and Marburg virus disease. It also shows broad-spectrum antiviral activity against RNA virus families including corona viruses.

Multiple other RdRp inhibitors are described in literature, for example broad spectrum antiviral 6’-bis fluorinated aristeromycin analogue (27). Chloroquine (28) is an existing anti malaria medicine. It blocked virus infection with an EC50 of 1.13 μM. Its mode-of-action is unclear. However, chloroquine inhibits endosomal acidification, and thus could stop the release of viral DNA into the cytoplasm. It is under assessment in more than 100 patients at over ten hospitals in Beijing and Guangdong province. Plans for an additional study in Hunan province are underway.

2.5. Other viral proteins

The role of the other SARS-CoV-2 nucleocapsid protein (N-protein) as drug discovery targets is less clear. However the crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain was just published and will give structural insight as potential drug target. It is rapidly detected by antibodies in serum, plasma, and peripheral blood and might therefore serve to develop specific diagnostics.

2.6. Host Targets

Using methods of machine learning-enabled scientific literature analysis the biotech company BENVOLENT proposed the AP2-associated protein kinase 1 (AAK1) as a host target to fight 2019-nCoV. AAK1 is a key enzyme of receptor-mediated endocytosis which is a major mechanism of most viruses to enter their host cells. Thus, they predict the approved (for rheumatoid arthritis) kinase inhibitor baricitinib 29 to reduce the ability of the virus to infect lung cells (Fig.8).
3. Outlook

The recent emergence of SARS-CoV-2 has put the world on alert. The rapid worldwide spread, the high human-to-human transmissibility, combined with the inability to contain the pandemic is causing an increasing death toll and also considerable paralysis of the world economy. The COVID-19 could decrease and disappear, or could be established worldwide in the human population and reoccur seasonally in the future with the danger of mutations. However, it is very likely that in the upcoming years we will see more outbreaks from corona and other viruses. The basic, translational and public health research communities have to prepare for this much better. The outbreak has emphasized the urgent need for renewed efforts to develop broad-spectrum antiviral agents to combat coronaviruses: on the positive side many new information of the virus biology, the spread was immediately shared; on the negative side many past opportunities to develop antivirals against coronaviruses were not taken despite a large number of promising approaches and compounds. The past decade has shown that coronavirus outbreaks are regularly reoccurring with more or less health impact on human and livestock. It remains to be hoped that the current pandemic will slow down and end as predicted in summer. Furthermore, it turns out that containment measures are not effective to avoid more severe spread. It remains to be seen if efficient and long-lasting immunity will develop in the infected population with regard to future outbreaks and that pharmacological measures can be rapidly developed to be able to treat severely sick people.
ACKNOWLEDGMENTS

Research in the Dömling laboratory is sponsored through ITN “Accelerated Early stage drug diScovery” (AEGIS, grant agreement No 675555), the National Institute of Health (NIH) (2R01GM097082-05), the European Lead Factory (IMI) (grant agreement number 115489), the Qatar National Research Foundation (NPRP6-065-3-012), and COFUNDs ALERT (grant agreement No 665250) and Prominent (grant agreement No 754425) and KWF Kankerbestrijding grant (grant agreement No 10504). L.G is greatful for a CSC stipend. Both authors are grateful to Micky Tortorella (GBH) for helpful discussions.

References