Drug Repurposing for Candidate SARS-CoV-2 Papain-like protease (PLpro) Inhibitors by a combined in Silico Method

Milan Sencanski1, Vladimir Perovic1, Jelena Milicevic1, Tamara Todorovic2, Radivoje Prodanovic3, Veljko Veljkovic3, Slobodan Paessler4, Sanja Glisic1

1 Laboratory of Bioinformatics and Computational Chemistry, Institute of Nuclear Sciences Vinca, National Institute of the Republic of Serbia, University of Belgrade, Belgrade, Serbia; sencanski@vinca.rs (M.S.), vladaper@vinca.rs (V.P.), jdjordjevic@vinca.rs (J.M.), sanja@vinca.rs (S.G)
2 Faculty of Chemistry, University of Belgrade, Belgrade, Serbia; tamarat@chem.bg.ac.rs (T.T.), rprodano@chem.bg.ac.rs (R.P.)
3 Biomed Protection, Galveston, TX, 77550, USA; veljko@biomedprotection.com
4 Department of Pathology, University of Texas Medical Branch, Galveston, TX, United States
5 Institute for Human Infections and Immunity, University of Texas Medical Branch, Galveston, TX, United States; slpaessl@utmb.edu

* Correspondence: sanja@vinca.rs

Abstract: The need for an effective drug against COVID-19, is, after almost 18 months since the global pandemic outburst, still very high. A very quick and safe approach to counteract COVID-19 is in silico drug repurposing. The SARS-CoV-2 PLpro promotes viral replication and modulates the host immune system, resulting in inhibition of the host antiviral innate immune response, and therefore is an attractive drug target. In this study, we used a combined in silico virtual screening candidates for SARS-CoV-2 PLpro protease inhibitors. We used the Informational spectrum method applied for Small Molecules for searching the Drugbank database and further followed by molecular docking. After in silico screening of drug space, we identified 44 drugs as potential SARS-CoV-2 PLpro inhibitors that we propose for further experimental testing.

Keywords: SARS-CoV-2; Papain-like protease; drug repurposing; in silico; ISM; molecular docking

1. Introduction

Long time frames, high costs, and high failure rates are associated with the traditional drug discovery process. In the current pandemic, in which finding the right drugs is of the highest priority, drug repurposing is a tool of great help in the fight against the COVID-19 virus. FDA-approved drugs with known safety and clinical profile allow for the reposition of drugs in the treatment of COVID-19. Despite the challenges that accompany drug repurposing, the ability to identify new uses of old drugs in a relatively short time is a significant incentive to focus on drug reposition for COVID-19 [1]. In the search for anti-SARS-CoV-2 molecules, various computational methods were applied combined with experimental approaches, and thousands of articles were published [2].

SARS-CoV-2 has a single-strand, positive-sense RNA genome 1 with at least ten open reading frames (ORFs) (Malik YA, 2020). The largest ORF1ab encompassing around two-thirds of the virus genome encodes two large overlapping polyproteins, the pp1a and pp1ab, essential for viral replication and transcription, which go through proteolytic cleavage, generating 16 non-structural proteins (NSP). [3–6]

The processing of two large viral polyproteins is autocatalytic proteolysis processed by virally encoded cysteine proteases. Papain-like protease (PLpro), encoded by NSP3, recognizes the LXGG tetrapeptide motif found in-between viral proteins nsp1 and nsp2, nsp2 and nsp3, and processes the replicase polyprotein 1a (pp1a) and replicase polyprotein 1ab (pp1ab) on the N-termini into nsp1, nsp2, and nsp3, essential for viral replication [7,8]. PLP is a monomer with an active site that comprises a Cys111/His272/Asp239 catalytic triad. The SARS-CoV-2 PLpro modulates the host immune system via deubiquitination and deISGylation from the host cell proteins resulting in inhibition of the host antiviral innate immune response [9,10].

In silico strategy of repurposing approved drugs has been employed to fight COVID-19 in the current study. We used a virtual screening protocol with combined sequential filters based on long-range and short-
range interactions to select candidates for PLpro inhibitors. The information spectrum method applied to small molecules (ISM-SM) was used for search the Drugbank database, followed by molecular docking. By using this combined protocol, 44 compounds were selected for further experimental testing.

2. Materials and Methods

2.1. Informational spectrum method

In this work, we analyze SARS-CoV-2 protein Papain like protease protein using the informational spectrum method (ISM). A comprehensive explanation of the sequence analysis based on ISM is available elsewhere [11]. According to this approach, sequence (protein or DNA) is transformed into a signal by assignment of numerical values of each element (amino acid or nucleotide). These values correspond to electron-ion interaction potential (EIIP) [12], determining the electronic properties of amino acid/nucleotides, which are essential for their intermolecular interactions. The EIIP descriptors are easily calculated using following formulas:

\[ Z^* = \sum_{i=1}^{m} n_i Z_i / N, \]

\[ EIIP = 0.25Z^* \sin(1.04\pi Z^*) / 2\pi, \]

Where \( i \) is type of the chemical element, \( Z \) is valence of the \( i \)-th chemical element, \( n \) is number of the \( i \)-th chemical element atoms in the compound, \( m \) is number of types of chemical elements in the compound and \( N \) is total number of atoms.

The EIIP signal is then transformed using Fast Fourier Transform (FFT) into information spectrum (IS) as a representation of a sequence in the form of a series of frequencies and amplitudes:

\[ X(n) = \sum_{m=1}^{N} x(m)e^{-i\pi mn / N}, n = 1, 2, ..., N/2. \]

Where \( m \) is the summation index, \( x(m) \) is the \( m \)-th member of a given numerical “signal” series (from a transformed, encoded primary protein sequence in our case), \( N \) is the total number of points in this series), \( n \) is the number of a discrete frequency (ranging from 1 on up to \( N/2 \)) in the DFT, \( X(n) \) are the discrete Fourier transformation amplitude coefficients corresponding to each discrete frequency \( n \), and \( 2\pi(n/N) \) is the phase angle at each given \( m \) in the amino-acid series of the protein in question.

However, in the case of protein analysis, the relevant information is primarily presented in energy density spectrum, which is defined as follows:

\[ S(n) = X(n)X^*(n) = |X(n)|^2, n = 1, 2, ..., N/2, \]

By this, the virtual spectroscopy method is feasible to analyze protein sequences without any previous experimental data functionally. Its extension for small molecules, ISM-SM was developed and published recently [13]. A small molecule is imported in smiles notation and decoded by atomic groups into an array of corresponding EIIP values. Using FFT, the corresponding IS of a small molecule is computed. This spectrum is further multiplied by IS of the protein receptor to obtain a Cross-spectrum (CS). Cross-spectral function is the function which determines common frequency characteristics of two signals. For discrete series it is defined as follows:

\[ S(n) = X(n) \ast Y(n)^*, n = 1, 2, ..., N/2, \]

Where \( X(n) \) and DFT coefficients of the series \( x(m) \), and \( Y(n)^* \) are complex conjugated DFT coefficients of the series \( Y(m) \).

From common frequencies in CS, one can determine whether protein interacts with small molecule and determine the corresponding binding region in the protein.

2.2. Data preparation

FASTA SARS-CoV-2 PLpro sequence was downloaded from UNIPROT and corresponding IS was calculated. A set of 1490 approved Drugbank[14] drugs with corresponding SMILES was subjected to IS and CS calculation with PLpro. All calculations were carried using our in-house software. PDB structures
of PLpro in complexes with inhibitors (in brackets), encoded 6WUU[15] (VIR250), 7CJM [16] (GRL0167), 7JIW [17] (VBY501) and 7LBR [18] (XR8-89) were downloaded from RCSB Protein Bank Database.

2.3. Molecular docking

Molecular docking of selected candidates into the crystal structure of PLpro was carried. Receptor three-dimensional structure was downloaded from RCSB, PDBID 7CJM[16]. All ligands, waters and ions were removed from PDB file. Two grid boxes with dimensions 24 x 24 x 24 Å were set to span all amino acid residues interacting with co-crystallised inhibitor GRL 0617. The (x,y,z) centers of the grid boxes was (26.0, 70.0, -1.0). Selected drugs from previous step were converted from SMILES to 3D SDF and further to PDB files and protonated at physiological pH. Geometry optimization was carried in MOPAC 2016[19] at PM7[20] level of theory. Default software settings for hydrophobic and hydrophilic terms in docking search function were used. Exhaustiveness was set to 50. Molecular docking was carried in Autodock Vina 1.1.2 [21].

Figures were made in BIOVIA Discovery Studio 2017, Schrodinger Maestro 11.1 and Origin 9.0 software.

3. Results

3.1. Informational spectrum method analysis

In the present study, we have used the Informational spectrum method (ISM) for the structure/function analysis of SARS-CoV-2 protein PLpro. The informational spectrum (IS) of PLpro contains two characteristic peaks, at the dominant frequencies F(0.383), and F(0.279), shown in Figure 1. To find the domains of a protein crucial for the information related to the three frequencies, PLpro was computationally scanned. As a result of scanning with the ISM algorithm, we identified regions with the highest amplitudes at these frequencies and shown that the regions, including residues 248-312 and 60-124, are essential for the information represented by the frequency F(0.383) and F(0.279), respectively. This finding is in accordance with the co-crystalized covalently bound peptide-like inhibitor VIR250[15] (Figure 2). Namely, those two dominant frequencies correspond to the inhibitor domain of the enzyme, mapping the regions of key amino acids and the binding site of the reported co-crystallized inhibitors. We further searched CS of Drugbank[14] candidates with PLpro at the F(0.383) and F(0.279) to find potential inhibitor candidates. With this search, we selected 44 candidate drugs (Table 1).
**Figure 1.** ISM spectrum of SARS-CoV-2 Papain like protease

**Figure 2.** Crystal structure of PLpro, with marked catalytic triad (PDBID 6WUU). Marked regions correspond F(0.383) 248-312 – blue, and F(0.279) 60-124 – green. The bound compound is the co-crystallized covalent peptide inhibitor VIR250.

<table>
<thead>
<tr>
<th>Compound name</th>
<th>Drugbank ID</th>
<th>F</th>
<th>VINA binding energy (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIR250</td>
<td>-</td>
<td>0.279</td>
<td>-</td>
</tr>
<tr>
<td>GRL 0617</td>
<td>-</td>
<td>0.382</td>
<td>-10.4</td>
</tr>
<tr>
<td>Epicriptine</td>
<td>DB11275</td>
<td>0.382</td>
<td>-10.3</td>
</tr>
<tr>
<td>XR8-89</td>
<td>-</td>
<td>0.279</td>
<td>-10.2</td>
</tr>
<tr>
<td>VBY501</td>
<td>-</td>
<td>0.382</td>
<td>-10.2</td>
</tr>
<tr>
<td>Metergoline</td>
<td>DB13520</td>
<td>0.382</td>
<td>-9.4</td>
</tr>
<tr>
<td>Dihydro-alpha-ergocryptine</td>
<td>DB11274</td>
<td>0.382</td>
<td>-8.9</td>
</tr>
<tr>
<td>Digoxin</td>
<td>DB00390</td>
<td>0.279</td>
<td>-8.8</td>
</tr>
<tr>
<td>Phentolamine</td>
<td>DB00692</td>
<td>0.279</td>
<td>-8.3</td>
</tr>
<tr>
<td>Fidaxomicin</td>
<td>DB08874</td>
<td>0.382</td>
<td>-8.1</td>
</tr>
<tr>
<td>Ergometrine</td>
<td>DB01253</td>
<td>0.382</td>
<td>-8.0</td>
</tr>
<tr>
<td>Sacubitril</td>
<td>DB09292</td>
<td>0.279</td>
<td>-7.8</td>
</tr>
<tr>
<td>Almitrine</td>
<td>DB01430</td>
<td>0.382</td>
<td>-7.7</td>
</tr>
<tr>
<td>Osimertinib</td>
<td>DB09330</td>
<td>0.279</td>
<td>-7.6</td>
</tr>
<tr>
<td>Indacaterol</td>
<td>DB05039</td>
<td>0.279</td>
<td>-7.6</td>
</tr>
<tr>
<td>Methylergometrine</td>
<td>DB00353</td>
<td>0.382</td>
<td>-7.6</td>
</tr>
<tr>
<td>Diacetyl benzoyl lathyrol</td>
<td>DB11260</td>
<td>0.279</td>
<td>-7.5</td>
</tr>
<tr>
<td>Bosutinib</td>
<td>DB06616</td>
<td>0.279</td>
<td>-7.4</td>
</tr>
<tr>
<td>Tubocurarine</td>
<td>DB01199</td>
<td>0.279</td>
<td>-7.4</td>
</tr>
<tr>
<td>Terconazole</td>
<td>DB00251</td>
<td>0.382</td>
<td>-7.3</td>
</tr>
<tr>
<td>Ixabradine</td>
<td>DB09083</td>
<td>0.382</td>
<td>-7.0</td>
</tr>
<tr>
<td>Vandetanib</td>
<td>DB05294</td>
<td>0.382</td>
<td>-7.0</td>
</tr>
<tr>
<td>Citalopram</td>
<td>DB00215</td>
<td>0.382</td>
<td>-7.0</td>
</tr>
</tbody>
</table>
### Table 2. Interaction scheme between PLpro inhibitor binding site aminoacid residues and ligands (X depicts the presence of interaction)

<table>
<thead>
<tr>
<th>Aminoacid residue</th>
<th>GRL 0617</th>
<th>Epicriptine</th>
<th>Methylergoline</th>
</tr>
</thead>
</table>

3.2. Molecular docking

To further filter the selected compounds, we carried molecular docking into the site of reported co-crystallized PLpro inhibitors, using PLpro - GRL 0617 complex structure (PDBID 7CJM) [16]. The binding energies values were compared to the docked co-crystallized ligands from different PLpro structures. From the initial docking, as top candidates were found Epicriptine and Methylergoline, targeting the inhibitor binding site. Epicriptine is the top candidate, with a more favourable docking energy (-10.3 kcal/mol) than literature inhibitors XR8-89[18] and VBY501[17], suggesting it could be a potentially promising inhibitor of SARS-CoV-2 PLpro. Methylergoline is the second best candidate, binding to the PLpro slightly weaker than literature inhibitor VBY501 (-9.4 kcal/mol). Types of intermolecular interactions that candidates form with amino acid residues are hydrogen bonds, aromatic π-π, alkyl-π, cation-π interactions. Comparing the binding patterns of our top candidates to the literature inhibitors, it is noticeable that they are conserved. Thus, both Epicriptine (Figure 3) and Methylergoline (Figure 4) form a salt bridge with Asp164, analogously to GRL 0617 (Figure 5). Methylergolin, in common with GRL 0617, forms hydrogen bond with Gln269 via carboxyl oxygen. In all three cases, the aromatic moieties are oriented towards Tyr 268. In addition, Epicriptine forms cation-π interaction with Tyr 268. The summary of protein-ligand intermolecular interactions is presented in Table 2.
<table>
<thead>
<tr>
<th>Residue</th>
<th>ASP164</th>
<th>X</th>
<th>X</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residue</td>
<td>GLN269</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Residue</td>
<td>GLU167</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residue</td>
<td>GLY163</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residue</td>
<td>LEU162</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residue</td>
<td>MET208</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residue</td>
<td>PRO247</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Residue</td>
<td>PRO248</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Residue</td>
<td>TYR264</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Residue</td>
<td>TYR268</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residue</td>
<td>TYR273</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 3.** Epicriptine in the PLpro inhibitor binding site. Green lines: hydrogen bonds; orange: electrostatic interactions; purple: alkyl-π interactions, magenta: π-π interactions
Figure 4. Metergoline in the PLPro inhibitor binding site. Green lines: hydrogen bonds; orange: electrostatic interactions; purple: alkyl-π interactions, magenta: π-π interactions
Figure 5. Co crystalized ligand GRL 0617 in the PLpro allosteric site. Green lines: hydrogen bonds; purple: alkyl-π/hydrophobic interactions.

4. Discussion

The fundamental problem in a search for preventive and therapeutic options to respond to threats of a pandemic is a costly, time-consuming, and risky process of drug development. As an antiviral agent, remdesivir is currently the only drug fully approved for the treatment of COVID-19, there is an urgent need for efficient antivirals against SARS-CoV-2 infection [22].

A promising therapeutic strategy for many viral diseases and the most rational in the current pandemic is drug repurposing. Given that the in silico approach allows rapid screening of large collections of compounds, computational drug repurposing offers a promising route when time is a critical factor. Various computational predictions approaches have been developed to identify drug repositioning opportunities against SARS-CoV-2 [23,24].

However, since it is challenging to simulate complex biological structures computationally, using state-of-the-art methods have advantages and limitations. It has been shown that the use of combined in silico approaches provide strong grounds for repurposed hypothesis worthy of experimental investigation. [25]

In this manuscript, the original concept in selecting candidate molecules for the treatment of SARS-CoV-2 infection was applied. The concept includes molecular characteristics responsible for long-term recognition between biological molecules in interaction. The VS protocol in this study is based on a combined in silico approach, which implies both short- and long-range interactions between interacting molecules.
The ISM was used in this work for the structure/function analysis of the SARS-CoV-2 protein PLpro and identification of the main informational characteristic of the protein, which corresponds to the protein key biological function. At the beginning of the pandemic, the use of the ISM was for the first time suggested a potential SARS-CoV-2 receptor, therapeutic / vaccine target, and proposed SARS-CoV-2 cell to cell transmission.[26] In another recent study, ISM was used for analyzing SARS-CoV-2 Orf3b, suggesting that this protein acts as a modulator of the interferon signaling network.[27]

Previously, by using the same combined VS protocol as in this study, we have selected candidate SARS-CoV-2 Mpro inhibitors and proposed 57 compounds for further experimental testing [28]. The same approach has been successfully applied in analyzing GPCR drugs from the Golden dataset [13].

Anti SARS-CoV-2 activity was previously shown in other studies for some of the candidates PLpro inhibitors from the current study, however, against a different target or without a known mechanism of action. The potential multitarget activity of drugs proposed as PLpro inhibitors may help overcome drug resistance in COVID-19.

Using bioinformatics analysis and experimentally, it was shown that Methergoline, the top-ranked candidate for PLpro inhibitor according to our study, prevents SARS-CoV-2 infection primarily by interfering with viral entry [29]. One of the best ranked PLpro inhibitors from our computational study is digoxin. In an in-vitro study, digoxin leads to SARS-CoV-2 inhibition at the post-entry stage of the viral life cycle[30]. Digoxin is a cardiovascular drug with antiviral activity against several coronaviruses and is proposed as a potential COVID-19 therapeutic, with possible additional therapeutic effects for patients with cardiovascular disease. [30]

Osimertinib, a promising PLpro catalytic site inhibitor candidate according to our study, was identified as an inhibitor for spike-mediated entry that showed greater than 50% rescue of the SARS-CoV-2 cytopathic effect [31]. Recently in a high throughput screening assay for SARS-CoV-2, bosutinib was identified as a specific SARS2-S pseudovirus entry inhibitor that significantly inhibited SARS2 replication [32]. In our analysis, several clinically used antidepressants, including citalopram, proposed as PLpro inhibitor, showed antiviral action against SARS-CoV-2. [33]

Quinine, proposed as PLpro inhibitor in this work, inhibited SARS-CoV-2 infection and the toxicological and pharmacological profile seems more favorable when compared to its progeny drugs hydroxychloroquine or chloroquine[34]. It is also suggested that quinine with a reputation record as medication against feverish illnesses might be able to mitigate the cytokine storm associated with severe COVID-19 [34]. Methdilazine was reported for activity against SARS-CoV-2 infection and was among the top confirmed anti-SARS-CoV-2 compounds and was suggested as a PLpro inhibitor in this study [35].

For some of the compounds like fidaxomicin, proposed in our study as PLpro inhibitor, it has been previously proposed their antiviral activity from virtual screens against other SARS-CoV-2 targets, but without supporting biological data [36].

5. Conclusion

The ISM-SM approach has the advantage compared to other in silico approaches for its capability to determine long-range molecular recognition and targeting between protein and ligand. This approach determines with immense precision the location of the protein domains with the possible binding site and enables the selection of small molecules which have great specificity for proposed domains. Moreover, rapid scanning of large compound libraries is undemanding since it is needed only protein sequence and SMILES molecules notation for data preparation. PLpro inhibitor candidates proposed in our computational study should be further experimentally tested in searching for safe, effective new treatments against SARS-CoV-2.

Author Contributions: Conceptualization, M.S., S.G., V.V. and S.P.; investigation, M.S., V.P., V.V. and S.G.; formal analysis, M.S., V.P., S.P., V.V. and S.G.; data curation, M.S., V.V. and S.G.; project administration, S.G.; writing—original draft preparation, M.S., V.P., V.V., R.P., T.T. J.M., S.P. and S.G.; writing—review and editing, M.S., V.P., V.V., R.P., T.T. J.M., S.P. and S.G.; software, M.S. and V.P.; visualisation, M.S. and V.P.; methodology, M.S. and V.P. All authors have read and agreed to the published version of the manuscript.
Funding: This work was supported by the Science Fund of the Republic of Serbia, as part of the Special Research Program on COVID 19 – the project COVIDTARGET – Repurposing of drugs for prevention and treatment of COVID-19

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References


22. FDA FDA Approves First Treatment for COVID-19; 2020;


