Screening potential anti-virals for the main protease of the Coronaviridae family including SARS-CoV-2, SARS-CoV, MERS

Clifford W Fong

Screening potential anti-virals for the main protease of the Coronaviridae family including SARS-CoV-2, SARS-CoV, MERS

Clifford W. Fong

Eigenenergy, Adelaide, South Australia, Australia.
Email: cwfong@internode.on.net

Keywords: COVID-2019 or SARS-CoV-2; SARS-CoV; MERS; 3C-like protease, or 3CLpro, or M^{pro}; inhibition; host cell membrane transport, Coronaviridae family, linear free energy relationships, HOMO-LUMO; quantum mechanics;

Abbreviations: Structure activity relationships SAR, $\Delta G_{desolv,CDS}$ free energy of water desolvation, $\Delta G_{lipo,CDS}$ lipophilicity free energy, CDS cavity dispersion solvent structure of the first solvation shell, Dipole moment DM, Molecular Volume Vol, HOMO highest occupied molecular orbital, LUMO lowest unoccupied molecular orbital, HOMO-LUMO energy gap, linear free energy relationships LFER

Abstract

It has been shown that structural activity of series of inhibitors based on a linear free energy relationship involving the water desolvation, lipophilicity, dipole moment and the HOMO-LUMO, LUMO or HOMO applies to the bat HKU4 $3CL^{pro}$ and SARS-CoV $3CL^{pro}$. Since the SARS-CoV and SARS-CoV-2 $3CL^{pro}$ have almost 100% identity similarity, these LFERs are highly likely to apply the SARS-CoV-2 $3CL^{pro}$. Also as the bat HKU4 $3CL^{pro}$ shares high sequence identity (81%) with the MERS-CoV protease these LFERs may also apply to MERS.

Previously it has been demonstrated that some anti-virals, viz Boceprevir, Telaprevir and the 13b α-ketoamide, show high inhibitory activity against the $3CL^{pro}$ of many members of the Coronaviridae family including those from bats, civets, cats and birds. It is suggested that the molecular specifiers used to derive LFERs for inhibition of coronavirus proteases can be used to predict which anti-virals will have high inhibitory capability as well as high host cell membrane transport capability against the $3CL^{pro}$ main protease of members of the Coronaviridae family.

This work shows that it may be possible to develop quantitative LFERs which may have applications to current and future coronaviruses that may cause human respiratory infections. Favipiravir, GS441524, Saquinavir, Ledipasvir, Nitazoxanide, Ruxolitinib, Baricitinib, Carfilzomib, Efavirenz, and 13b α-ketoamide are predicted to be the best inhibitors.

Introduction

The current pandemic caused by the SARS-CoV-2 corona virus has created enormous effort to find a vaccine against this virus, as well as the search for anti-viral therapies which can ameliorate the symptoms of the infection. However the search for an vaccine is an extensive process and previous experience in finding vaccines against corona viruses that particularly affect the upper respiratory tract has not been rewarded. The search for therapeutic anti-virals,

particularly those repurposed from other viral infections, is a less onerous task, but less clinically effective than finding a vaccine.

The history of SARS-CoV, MERS, and now SARS-CoV-2, suggest that there will be future corona viruses that will cause human epidemics or pandemics. [1,2] [Fani, Rabaan] Treating hypothetical future viruses with vaccines is not feasible, but there exists the possibility that current anti-virals that are effective against SARS-CoV, MERS, and SARS-CoV-2 could also be effective against future corona viruses. All coronaviruses require the proteolytic activity of nsp5 protease (aka *3C-like protease, 3CL^{pro} or M^{pro}*) during virus replication, making it a high value target for the development of anti-coronavirus therapeutics. The protease enzymes from different CoV strains are known to share high sequence and 3D structure homology providing a strong structural rationale for designing wide-spectrum anti-CoV inhibitors.

Ortega [3] has shown that the main protease of SARS-CoV-2 is highly conserved (96-99%) among all the SARS-like coronaviruses from human and related CoVs from *other animals particularly bats and civets*. However the MERS-CoV protease was only 50.6% identical. The SARS-CoV and SARS-CoV-2 main proteases exhibit 96 % sequence identity to each other.

A comparison of 49 structures of M^{pro} from different coronavirus strains from the Coronaviridae family showed the main protease structural backbone and active site conformation are conserved despite sequence variations. Substrate recognition and binding efficiency or function of the M^{pro} are highly conserved across the virus family. Protease structures from the coronavirus strains causing human respiratory infections like SARS, MERS and SARS-CoV-2 as well as those from bats were highly conserved. The phylogenetic analysis validated the bat origin of the SARS-CoV-2. [4] [Joshi]

The similarities amongst the M^{pro} of the Coronaviridae family indicate that the development of universal M^{pro} assays and the design of broad-spectrum inhibitors of M^{pro} from different coronavirus strains are highly prospective. [3-6] [Ortega, Joshi, Wang, Yang] Further the prospects for therapeutically treating future corona viruses that cause human respiratory infections may be enhanced by recognizing that anti-virals effective against SARS-CoV, MERS, SARS-CoV-2, and other bat or civet viruses may be effective against future similar corona viruses that may evolve.

Anson [7] has found the FDA approved drugs Boceprevir and Telaprevir (see Figure 1) showed significant inhibition of 3CL^{pro} from SARS-CoV-2 with IC_{50} values of 2.5 and 10.7 μM respectively. Broad spectrum inhibition of the potency of Boceprevir and Telaprevir against a panel of eight purified 3CL^{pro} enzymes from the Alpha-, Beta-, and Gamma-coronaviral proteases, including the human coronaviruses SARS, MERS, NL63, HKU1, bat coronaviruses HKU4 and HKU5, and the animal coronaviruses FIPV (cat) and IBV (bird) has been tested. Boceprevir inhibits eight out of the nine 3CL^{pro} enzymes tested, with IC_{50} values ranging from 420 nM to 9.2 μM, while Telaprevir inhibits six out of the nine 3CL^{pro} enzymes with IC_{50} values ranging from 2.4 μM to 35 μM. Boceprevir exhibited equipotent inhibition of SARS-CoV-2 and SARS-CoV 3CL^{pro} (IC_{50} ~2.6 μM), which was expected given that their active sites are 96% identical. It is noted that while Telaprevir did inhibit the enzymes from bat HKU4 and HKU5 coronaviruses, it exhibited little to no inhibition of the MERS-CoV 3CL^{pro}. However Zhang [8]

found that the α-ketoamide 13b did inhibit the MERS-CoV $3CL^{pro}$. The MERS-CoV $3CL^{pro}$ is unique among coronavirus proteases in exhibiting substrate-induced dimerization because it exists as a weakly associated dimer, whereas SARS-CoV, HKU4 and HKU5 $3CL^{pro}$s on the other hand are tightly associated dimers. [7,9] [Anson, Tomar]

Zhang [8] studied some α-ketoamide compounds and found that compound 13b (see Figure 1) inhibited the purified recombinant SARS-CoV-2 M^{pro} with IC_{50} = 0.67 μM, whereas the corresponding IC_{50} values for inhibition of the SARS-CoV M^{pro} and the MERS-CoV M^{pro} were 0.90 and 0.58 μM respectively. These IC_{50} values are equivalent within experimental error, so these data and other studies [3-6] [Joshi, Ortega, Wang, Yang] indicate that the protease has a very close identity in the SARS-CoV-2, SARS-CoV and MERS-CoV replicating process.

We have previously used the inhibition of the bat HKU4 coronavirus protease to derive a quantum mechanical linear free energy (LFER) structure activity relationship between pIC_{50} and $\Delta G_{desolv,CDS}$ the free energy of water desolvation, $\Delta G_{lipo,CDS}$ the lipophilicity free energy, the dipole moment in water, and HOMO-LUMO the energy gap in water of the inhibitors. [10] [Fong 2020] Also the transportation of anti-virals into the infected host cell has been shown to be an equally important factor in determining the anti-viral therapeutic efficacy as the inhibition of the coronavirus proteases itself. Small anti-virals that are neutral at physiological pH have the best chance of being transported across the cell membrane. Larger anti-virals which may use endocytosis to enter the host cell will require some degree of desolvation of the drug before being engulfed by the lipophilic membrane, and by doing so, lower the energy required to promote endocytosis. [11] [Fong 2020]

HKU4-CoV $3CL^{pro}$ shares high sequence identity (81%) with the MERS-CoV enzyme and has been shown to be a surrogate model for evaluating anti-virals against MERS, [12] [Abuhammad 2017] while Anson [7] and Zhang [8] have shown that some inhibitors are effective inhibitors of SARS-Cov-2, SARS-CoV, HKU4 and sometimes against MERS. A comparison of the X-ray crystal structures of the bat HKU4-CoV $3CL^{pro}$ with the SARS-CoV-2 $3CL^{pro}$ reveals a 65% sequence similarity. [13] [Stoermer]

It is known that the M^{pro} in SARS-CoV and SARS-CoV-2 have a very high similarity 96% identity sequence, including 100% active site conservation. Bobrowski [14] evaluated a large number of anti-viral potentially active against the M^{pro} of SARS-Cov-2 using molecular docking and structure activity studies, but found that molecular docking could *not* discriminate between experimentally active and inactive inhibitors, whereas structure activity relationships were better predictors of efficacy. This molecular docking observation is interesting since docking is now the main technique to evaluate the efficacy of potential protease inhibitors. [15] [Ton]

This paper aims to:

(a) Examine the structural activity LFER of series of inhibitors of the SARS-CoV $3CL^{pro}$ to compare with the previous LFER derived for the bat HKU4 $3CL^{pro}$ and so test the generality of the LFER method, and
(b) Test the ability of the derived LFERs to screen repurposed anti-virals for their efficacy against SARS-CoV-2, and

(c) Examine several anti-virals Boceprevir, Telaprevir and the 13b α-ketoamide which have been shown to have wide protease inhibitory activity for many members of the Coronaviridae family, including bats, civets, cats and birds, and which may have applications to future coronaviruses that may cause human respiratory infections

Results and evaluation methodology

We have recently developed a quantum mechanically based structure activity model for the inhibition of the 3C-like protease from the bat coronavirus HKU4 (HKU4-CoV) [10] [Fong 2020] for a particular series of inhibitors. HKU4 (HKU4-CoV) belongs to the same 2c lineage as MERS-CoV and shows high sequence similarity with MERS-CoV. HKU4-CoV 3CLpro shares high sequence identity (81%) with the MERS-CoV enzyme and has been shown to be a surrogate model for evaluating anti-virals against MERS. [12] [Abuhammad 2017]

We have previously shown that equation 1 accurately describes the inhibition of the HKU4-CoV $3CL^{pro}$:

Eq 1 Inhibition of 3C-like protease pIC_{50} for 40 compounds from [16] [Lin] was:

$$\mathbf{pIC_{50} = 0.05\Delta G_{desolv,CDS} - 0.11\Delta G_{lipo,CDS} - 0.08Dipole\ Moment - 0.23(HOMO\text{-}LUMO) + 6.63}$$

Where R^2 = 0.382, SEE = 0.39, SE($\Delta G_{desolvCDS}$) = 0.04, SE($\Delta G_{lipoCDS}$) = 0.04, SE(Dipole Moment) = 0.025, SE(HOMO-LUMO) = 0.08, F=5.42, Significance= 0.0017

where $\Delta G_{desolv,CDS}$ is the free energy of water desolvation, $\Delta G_{lipo,CDS}$ is the lipophilicity free energy, the dipole moment in water, and HOMO-LUMO is the energy gap in water.

The important finding was that pIC_{50} is dominantly related to the HOMO-LUMO energy gap of the inhibitors. The HOMO-LUMO gap is an inherent descriptor of the innate reactivity of the inhibitor, and is related to how the inhibitor binds to the protease. In particular, how the HOMO of the protease ($HOMO_{prot}$) interacts with the LUMO of the inhibitor ($LUMO_{inhib}$), and how the HOMO of the inhibitor ($HOMO_{inhib}$) interacts with the LUMO of the protease ($LUMO_{prot}$). These molecular interactions fundamentally define the inhibitor-protease binding interaction.

The evaluation method used in this study and others was to calculate the *molecular specifiers* for the various inhibitors used on the 3C-like protease: (1) the free energy of water desolvation ($\Delta G_{desolv,CDS}$), (2) the lipophilicity free energy ($\Delta G_{lipo,CDS}$) in n-octane, (3) the dipole moment in water, (4) the molecular volume in water, and (5) HOMO, (6) LUMO or (7) HOMO-LUMO energy gap in water. These independent variables values can be scaled to similar magnitudes so that the coefficients in the multiple linear regression equations can be directly compared to gauge the relative magnitudes of inhibitory sensitivity of these molecular variables. Stepwise multiple regression is then applied to seek out which of the seven drug molecular properties had the largest and most significant effect on the inhibition. Equation 1 shows the most statistically significant relationship found after testing against all independent variables in a stepwise fashion.

In this study we have used the same methodology to evaluate the inhibition of the SARS-CoV 3CLpro by two series of inhibitors, shown in Table 1 and Figure 2. It is known that the M^{pro} in

SARS-CoV and SARS-CoV-2 have a very high similarity 96% identity sequence, including 100% active site conservation.

Wang [17] studied the inhibitory behaviour of a range of unsymmetrical aromatic disulphides against the SARS M^{pro} which covered the IC_{50} range of 0.5 to 6 µM. (see Table 1 and Figure 2) Wang utilized comparative field analysis to derive SAR and docking studies. These data are highly accurate IC_{50} values and cover a wide range of 40 structures, and the strong correlation found is shown in eq 2.

Eq 2(a) Inhibition of 3C-like protease of SARS IC_{50} for 35 compounds (excluding 5 outliers) from [17] [Wang] was:

$$\mathbf{IC_{50} = -0.74\Delta G_{desolv,CDS} - 0.30\Delta G_{lipo,CDS} - 0.21Dipole\ Moment + 0.25(HOMO\text{-}LUMO) - 2.31}$$

Where R^2 = 0.725, SEE = 0.69, SE($\Delta G_{desolvCDS}$) = 0.11, SE($\Delta G_{lipoCDS}$) = 0.075, SE(Dipole Moment) = 0.07, SE(HOMO-LUMO) = 0.38, F=19.75, Significance=0.000000

A slightly better correlation was found using eq 2(b) where the LUMO was substituted for the HOMO-LUMO energy gap, with the standard error of the LUMO variable having a P-value of 0.104 compared to the SE of the HOMO-LUMO variable having a P-value of 0.516.

Eq 2(b)

$$\mathbf{IC_{50} = -0.78\Delta G_{desolv,CDS} - 0.25\Delta G_{lipo,CDS} - 0.21Dipole\ Moment + 0.63LUMO - 0.272}$$

Where R^2 = 0.745, SEE = 0.666, SE($\Delta G_{desolvCDS}$) = 0.11, SE($\Delta G_{lipoCDS}$) = 0.08, SE(Dipole Moment) = 0.06, SE(LUMO) = 0.38, F=21.84, Significance=0.0000000

The corresponding correlation using the HOMO variable instead of the HOMO-LUMO variable is eq 2(c):

Eq 2(c)

$$\mathbf{IC_{50} = -0.78\Delta G_{desolv,CDS} - 0.25\Delta G_{lipo,CDS} - 0.255Dipole\ Moment + 0.68HOMO + 3.73}$$

Where R^2 = 0.737, SEE = 0.677, SE($\Delta G_{desolvCDS}$) = 0.11, SE($\Delta G_{lipoCDS}$) = 0.08, SE(Dipole Moment) = 0.065, SE(HOMO) = 0.52, F=20.86, Significance=0.0000000

Tsai [18] screened 59363 compounds by a structure-based virtual screening approach to identify inhibitors of SARS-CoV $3CL_{pro}$, and then tested the inhibition of 28 identified widely diverse compounds, (see Table 1 and Figure 2) which covered a IC_{50} range of 3-1000 µM. Unfortunately the experimental inaccuracy of the less active inhibitors was large, probably +/- 5 above an IC_{50} of 30 µM, and +/- 50 and even more at higher IC_{50} values over 60 µM. However the approximate relationships derived using these data can be usefully compared with the more accurate data from Wang, albeit from a different structural class of inhibitors.

Analysis of the 25 inhibitors of the SARS-CoV 3C-like protease investigated by Tsai [18] yields eq 3(a) and 3(b):

Eq 3(a)

$$\mathbf{IC_{50} = -25.2\Delta G_{desolv,CDS} + 19.7\Delta G_{lipo,CDS} + 5.3Dipole\ Moment + 24.1HOMO + 192.1}$$

Where R^2 = 0.347, SEE = 102.8, SE($\Delta G_{desolvCDS}$) = 10.6, SE($\Delta G_{lipoCDS}$) = 15.4, SE(Dipole Moment) = 7.8, SE(HOMO) = 41.8, F=2.79, Significance=0.053

Eq 3(b)

$$\mathbf{IC_{50} = -29.2\Delta G_{desolv,CDS} + 15.2\Delta G_{lipo,CDS} + 7.2Dipole\ Moment + 27.7LUMO + 4.3}$$

Where R^2 = 0.350, SEE = 102.5, SE($\Delta G_{desolvCDS}$) = 12.3, SE($\Delta G_{lipoCDS}$) = 15.5, SE(Dipole Moment) = 7.2, SE(HOMO) = 42.3, F=2.82, Significance=0.051

Eq 3(a) and 3(b) are low precision relationships due to the poor precision of the IC_{50} values, and as such can only be indicative rather than definitive. Two data points at the highest IC_{50} values ($IC_{50} > 350$) were clear outliers. The largest uncertainty lies with the Dipole Moment and HOMO or LUMO coefficients, as can be seen from the large SE's. No relationship was found with the HOMO-LUMO variable. Both equations eq 3(a) and 3(b) reinforce the much more precise eq 2(b) and 2(c) in validating that water desolvation, lipophilicity, dipole moment and either HOMO or LUMO are precise molecular specifiers that control the inhibition of the SARS-CoV 3C-like protease. We have tested the sensitivity of eq 3(a) or 3(b) to a change in the conformation of various anti-virals, with Table 1 showing for example, that compound 1 from Tsai [18] when tested in a different conformation (1 alt) has molecular specifiers that are virtually invariant to the change of conformation in water, other than, as expected, a change in the dipole moment.

We have previously used eq 1 which describes the inhibition of the HKU4-CoV $3CL^{pro}$ to rank the likely inhibition of the $3CL^{pro}$ of the SARS-CoV-2 by a wide range of currently available repurposed anti-virals. [11] [Fong] We have extended this study by using eq 2(b) and eq 3(a) for the inhibition of 3C-like protease of SARS for 35 disulphide compounds, and 25 other compounds to rank these repurposed anti-virals for potential efficacy against the $3CL^{pro}$ of the SARS-CoV-2. The 96% similarity of the $3CL^{pro}$ for SARS and SARS-Cov-2 makes this comparison more likely to be accurate that the lower 65% similarity of the HKU4-CoV $3CL^{pro}$ with SARS-Cov-2.

These results are shown in Table 2. These data can only relatively rank the likely efficacy of the various inhibitors to each other, since it is clear that these LFER equations are specific to a particular structural class of inhibitors active against various proteases, so the ranking derived from a particular LFER equation are dependent on the inhibitor class structures. We have previously shown that inhibitors that are predominantly charged at physiological pH levels will not easily passively permeate or be actively transported by endocytosis across the cell membrane of host cells, since desolvation of the inhibitors is greatly increased for charged inhibitors, as shown in Table 2. It should also be noted that the calculated inhibitory capability of the protonated and di-protonated anti-virals are less meaningful than those for the neutral species in Table 2, because equations 1, 2, and 3 were derived from various series of neutral anti-virals.

Discussion

It can be seen from eqs 1-3 that an equation of the general form is applicable to the inhibition of the SARS-CoV 3C-like protease and the HKU4-CoV 3C-like protease:

General form

$IC_{50} = \Delta G_{desolv,CDS} + \Delta G_{lipo,CDS}$ + Dipole Moment + LUMO (or HOMO or HOMO-LUMO)

This result is consistent with the previously known similarity of the 3C-like (or $3CL^{pro}$ or M^{pro}) protease of the SARS, HKU4-Cov and other members of the Coronaviridae family, which include those derived from bats, civets, birds and cats. Since the main protease from SARS-CoV has almost 100% identity similarity with that from the SARS-CoV-2 coronavirus, it is safe to conclude that an equation of the general form will be applicable to the inhibition of the SARS-CoV-2 3C-like protease.

The general form may also apply to the MERS virus, as the bat HKU4-CoV 3CLpro shares high sequence identity (81%) with the MERS-CoV enzyme and has been shown to be a good surrogate model for evaluating anti-virals against MERS. However, this connection may not be as pervasive as that for SARS-CoV since the MERS-CoV $3CL^{pro}$ is unique among coronavirus proteases in exhibiting substrate-induced dimerization because it exists as a weakly associated dimer, whereas SARS-CoV, HKU4 and HKU5 3CL proteases on the other hand are tightly associated dimers. [7,9] [Anson, Tomar]

We have earlier described how the HOMO-LUMO gap is an inherent descriptor of the innate reactivity of the inhibitor, and is related to how the inhibitor binds to the protease. In particular, how the HOMO of the protease ($HOMO_{prot}$) interacts with the LUMO of the inhibitor ($LUMO_{inhib}$), and how the HOMO of the inhibitor ($HOMO_{inhib}$) interacts with the LUMO of the protease ($LUMO_{prot}$). These molecular interactions fundamentally define the inhibitor-protease binding interaction, and can be used to predict the inhibitory binding of main proteases from a wide range of coronaviruses.

The work of Anson, Joshi and Zhang [4,7,8] has demonstrated that some anti-virals can inhibit the main protease form a wide range of viruses from Coronaviridae family. Table 1 also shows the molecular specifiers for the anti-virals Boceprevir, Telaprevir and the 13b α-ketoamide which have been shown to have wide protease activity of many members of the Coronaviridae family. Telaprevir shows a very small HOMO-LUMO gap of 1.71eV compared to values of 4.18 and 3.25 for Boceprevir and 13b α-ketoamide respectively, indicating that Telaprevir is much more reactive. However Telaprevir is also a much larger molecule 464 cm^3/mol compared to 374 and 392 cm^3/mol for Boceprevir and 13b α-ketoamide. Telaprevir also has a higher water desolvation penalty -19.65 kcal/mol compared to -13.13 and -13.37 kcal/mol for Boceprevir and 13b α-ketoamide. Since Boceprevir can inhibit the 3C-like protease from a wider range of coronaviruses than Telaprevir, these data suggest that while Telaprevir is more reactive than Boceprevir, its larger molecular size and larger desolvation requirements make it less potent in inhibiting the protease for some coronaviruses. The 13b α-ketoamide has similar molecular specifiers to that of Boceprevir, suggesting its inhibitory properties would be similar.

We have previously shown that in vivo, the inhibitory properties of anti-virals to treat coronaviruses will depend on how well the drugs can enter the infected host cells. [11] [Fong 2020] For diffusion dominant transport of drugs across the host cell membrane, small neutral anti-virals like favipiravir would have better membrane transport, as opposed to drugs that are charged at physiological pH levels. However for larger anti-virals, endocytosis is the likely

transport mechanism. Endocytosis requires active cellular energy input since substantial energy is required to allow the lipophilic cell membrane to engulf the drug and eventually deposit the drug into the cytoplasm. Endocytosis also requires some degree of desolvation of the drug before membrane engulfment, since the membrane is lipophilic, so desolvation, lipophilicity, and molecular size of the drug are major determinants of endocytosis.

Since the general form equation also applies to drug transport (as well as drug binding to enzymes) the molecular specifiers suggest that Boceprevir and the 13b α-ketoamide would be better transported across cell membranes, and would be better inhibitors of coronaviruses than Telaprevir.

The results in Table 2 for the repurposed inhibitors indicate that the general form of the LFER equation applies to the SARS-CoV, and bat HKU4 3CL proteases. The inhibitors colour coded in green in Table 2 show which inhibitors have high inhibitory capability while those colour coded in red are substantially protonated at the physiological pH, and therefore will not be transported easily across the cell membrane of infected cells. These anti-virals such as Hydroxychloroquine, Chloroquine, Nafamostst, Galidesivir, Remdesivir Triphosphate (shown as the neutral species, but almost certainly ionized at physiological pH) have high water desolvation penalties, compared to those neutral ant-virals coded in green. It is noteworthy that the same relative rankings of anti-virals with enhanced inhibitory capability (green colour coded) is shown for Favipiravir, GS441524, Saquinavir, Ledipasvir, Nitazoxanide, Ruxolitinib, Baricitinib, Carfilzomib, Efavirenz, and 13b α-ketoamide: ie the same relative ranking is obtained for the inhibition of the bat HKU4 $3CL^{pro}$ and SARS-CoV main proteases as calculated from eq 1 for the bat HKU4 3C-like protease, and eq 2(b) and 3(a) for the SARS-CoV 3C-like protease.

Table 2 also shows the predicted inhibitory capacity of Boceprevir, Telaprevir and the 13b α-ketoamide which have been previously shown to have wide activity protease activity of many members of the Coronaviridae family.

Conclusions

It has been shown that structural activity of series of inhibitors based on a linear free energy relationship involving the water desolvation, lipophilicity, dipole moment and the HOMO-LUMO, LUMO or HOMO applies to the bat HKU4 $3CL^{pro}$ and SARS-CoV $3CL^{pro}$. Since the SARS-CoV and SARS-CoV-2 $3CL^{pro}$ have almost 100% identity similarity, these LFERs are highly likely to apply the SARS-CoV-2 $3CL^{pro}$. Also as the bat HKU4 $3CL^{pro}$ shares high sequence identity (81%) with the MERS-CoV protease these LFERs may also apply to MERS.

Previously it has been demonstrated that some anti-virals, viz Boceprevir, Telaprevir and the 13b α-ketoamide, show high inhibitory activity against the $3CL^{pro}$ of many members of the Coronaviridae family including those from bats, civets, cats and birds. It is suggested that the molecular specifiers used to derive LFERs for inhibition of coronavirus proteases can be used to predict which anti-virals will have high inhibitory capability as well as high host cell membrane transport capability against the $3CL^{pro}$ main protease of members of the Coronaviridae family.

This work shows that it may be possible to develop quantitative LFERs which may have applications to current and future coronaviruses that may cause human respiratory infections.

Analysis of a series of repurposed anti-virals using the derived LFER equations for the bat HKU4 $3CL^{pro}$ and SARS-CoV $3CL^{pro}$ has predicted a number of anti-virals which can be effective against SARS-CoV-2 $3CL^{pro}$ as well as be effective in being transported across the cell membrane of coronavirus infected host cells. Favipiravir, GS441524, Saquinavir, Ledipasvir, Nitazoxanide, Ruxolitinib, Baricitinib, Carfilzomib, Efavirenz, and 13b α-ketoamide are predicted to be the best inhibitors.

Experimental Methods

All calculations were carried out using the Gaussian 09 package. Energy optimizations were at the DFT/B3LYP/6-31G(d) (6d, 7f) level of theory for all atoms in water. Selected optimizations at the DFT/B3LYP/6-311+G(d,p) (6d, 7f) level of theory gave very similar results to those at the lower level. Optimized structures were checked to ensure energy minima were located, with no imaginary frequencies. Energy calculations were conducted at the DFT/B3LYP/6-31G(d,p) (6d, 7f) for neutral and cationic compounds with optimized geometries in water, using the IEFPCM/SMD solvent model. With the 6-31G* basis set, the SMD model achieves mean unsigned errors of 0.6 - 1.0 kcal/mol in the solvation free energies of tested neutrals and mean unsigned errors of 4 kcal/mol on average for ions. [19] The 6-31G** basis set has been used to calculate absolute free energies of solvation and compare these data with experimental results for more than 500 neutral and charged compounds. The calculated values were in good agreement with experimental results across a wide range of compounds. [20,21] Adding diffuse functions to the 6-31G* basis set (ie 6-31+G**) had no significant effect on the solvation energies with a difference of less than 1% observed in solvents, which is within the literature error range for the IEFPCM/SMD solvent model. HOMO and LUMO calculations included both delocalized and localized orbitals (NBO). Experimental pKa's were available for Hydroxychloroquine Di-ion pKa 9.67 and 8.27 and Chloroquine Di-ion pKa 10.18 and 8.38 [22] Protonated anti-virals used pKa values were taken from Chemaxon: Nafamostat Di-ion, pKa 11.32, Galidesivir Ion pKa 8.46, Arbidol Ion pKa 9.87, Imatinib Ion pKa 8.27, Chloroquine Di-Ion pKa 10.32, Hydroxychloroquine Di-Ion pKa 9.76

It is noted that high computational accuracy for each species in different environments is not the focus of this study, but comparative differences between various species is the aim of the study, since the dominant errors in the LFER reside in the experimental IC_{50} or pIC_{50} values.

References

[1] M Fani, A Teimoori, S Ghafari, Comparison of the COVID-2019 (SARS-CoV-2) pathogenesis with SARS-CoV and MERS-CoV infections, Future Virology, 2020, 10.2217/fvl-2020-0050
[2] AA Rabaan, SH Al-Ahmed, S Haque, et al SARS-CoV-2, SARS-CoV, and MERS-CoV: a comparative overview, Le Infezioni in Medicina, 2020, 2, 174-184

[3] JT Ortega, ML Serrano, FH Pujol, H Rangel, Unrevealing sequence and structural features of novel coronavirus using in silico approaches: the main protease as molecular target, EXCLI Journal 2020,19, 400-409
[4] RS Joshi, SS Jagdale, SB Bansode, et al, Discovery of potential multi-target-directed ligands by targeting host-specific SARS-CoV-2 structurally conserved main protease, J Biomol Struct Dynam, 2020, DOI:0.1080/07391102.2020.1760137
[5] F Wang, C Chen, W Tan, K Yang, H Yang, Structure of main protease from human coronavirus NL63: insights for wide spectrum anti-coronavirus drug design. Sci Rep. 2016, 6, 22677 https://doi.org/10.1038/srep22677.
[6] H Yang, M Bartlam, Z Rao, Drug design targeting the main protease, the Achilles' heel of coronaviruses, Curr Pharm Des. 2006, 12, 4573-4590.
[7] BJ Anson, ME Chapman, EK Lendy, et al, Broad-spectrum inhibition of coronavirus main and papain-like proteases by HCV drugs, Researchsquare, 2020
[8] L Zhang, D Lin, X Sun, U Curth, C Drosten, L Sauerhering, S Becker, K Rox, R Hilgenfeld, Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved alpha-ketoamide inhibitors, Science, 2020, 368, 409–412.
[9] S Tomar, ML Johnston, SE St John, HL Osswald, PR Nyalapatla, LN Paul, AK Ghosh, MR Denison, AD Mesecar, Ligand-induced dimerization of MERS coronavirus nsp5 protease (3CLpro): implications for nsp5 regulation and the development of antivirals, J Biol Chem. 2015 Jun 8. pii: jbc.M115.651463. PMID:26055715
[10] CW Fong, Inhibition of COVID-2019 3C-like protease: structure activity relationship using quantum mechanics, 2020, hal-02529030v1
[11] CW Fong, Screening potential repurposed COVID-2019 3C-like protease inhibitors, 2020, hal-02663287v1
[12] A Abuhammad, RA Al-Aqtash, BJ Anson, AD Mesecar, MO Taha, Computational Modeling of the Bat HKU4 Coronavirus $3CL_{pro}$ Inhibitors as a Tool for the Development of Antivirals Against the Emerging Middle East Respiratory Syndrome (MERS) Coronavirus, J Mol Recognit 2017, 30(11):e2644, doi: 10.1002/jmr.2644. Epub 2017
[13] M Stoermer, Homology Models of Coronavirus 2019-nCoV 3CLpro Protease, 2020, ChemRxiv, Preprint, https://doi.org/10.26434/chemrxiv.11637294.v3
[14] T Bobrowski, VM Alves, CC Melo-Filho, D Korn, S Auerbach, C Schmitt, EN Muratov, A Tropsha, Computational models identify several FDA approved or experimental drugs as putative agents against SARS-CoV-2, ChemRxiv Preprint, https://doi.org/10.26434/chemrxiv.12153594.v1
[15] AT Ton, F Gentile, M Hsing, F Ban, A Cherkasov, Rapid Identification of Potential Inhibitors of SARS-CoV-2 Main Protease by Deep Docking of 1.3 Billion Compounds, Molec Inform 2020, DOI: 10.1002/minf.202000028
[16] F Lin, XM Fu, C Wang, SY Jiang, JH Wang, SW Zhang, L Yang, Y Li, QSAR, Molecular Docking and Molecular Dynamics of 3C-like Protease Inhibitors, Acta Phys. Chim. Sin. 2016, 32, 2693-2708
[17] L Wang, BB Bao, GQ Song, et al Discovery of unsymmetrical aromatic disulfides as novel inhibitors of SARS-CoV main protease: Chemical synthesis, biological evaluation, molecular docking and 3D-QSAR study, Eur J Med Chem 2017, 137, 450e461
[18] KC Tsai, SY Chen, PH Liang, et al, Discovery of a Novel Family of SARS-CoV Protease Inhibitors by Virtual Screening and 3D-QSAR Studies, J. Med. Chem. 2006, 49, 3485-3495

[19] AV Marenich, CJ Cramer, DJ Truhlar, Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions, J Phys Chem B, 2009, 113, 6378 -96
[20] S Rayne, K Forest, Accuracy of computational solvation free energies for neutral and ionic compounds: Dependence on level of theory and solvent model, Nature Proceedings, 2010, http://dx.doi.org/10.1038/npre.2010.4864.1.
[21] RC Rizzo, T Aynechi, DA Case, ID Kuntz, Estimation of Absolute Free Energies of Hydration Using Continuum Methods: Accuracy of Partial Charge Models and Optimization of Nonpolar Contributions, J Chem Theory Comput. 2006, 2, 128-139
[22] DC Warhurst, JCP Steele, et al Hydroxychloroquine is much less active than chloroquine against chloroquine-resistant *Plasmodium falciparum*, in agreement with its physicochemical properties, J Antimicrobial Chemotherapy, 2003, 52, 188–193

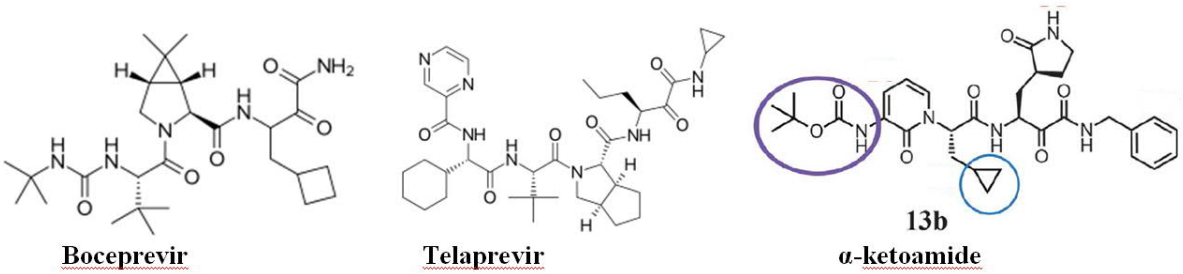

Figure 1. Structures of Boceprevir, Telaprevir, and the 13B α-ketoamide [8]

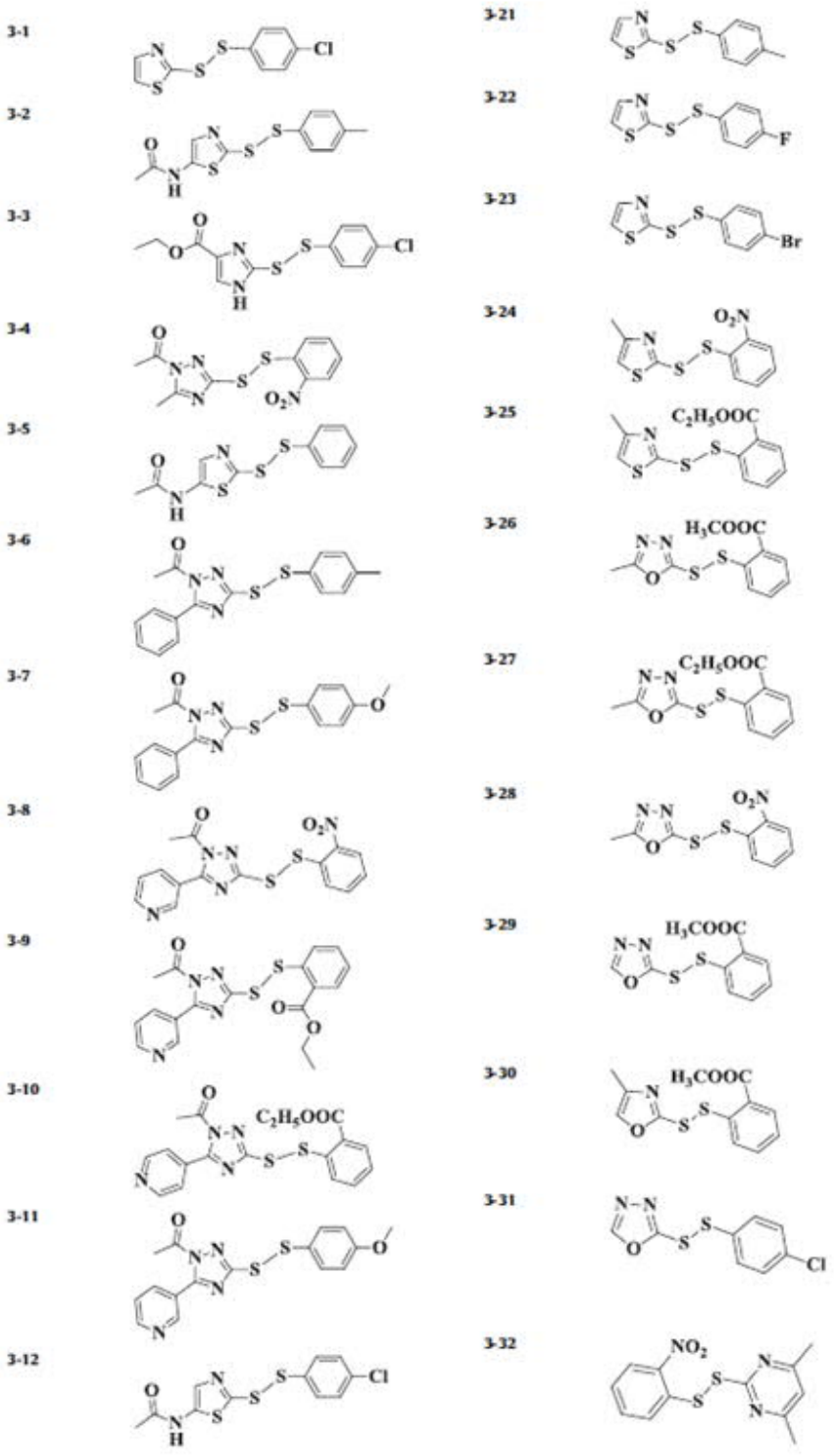

Figure 2(a)(i). Disulphide inhibitors of SARS-CoV 3CL proteases taken from Wang [17]

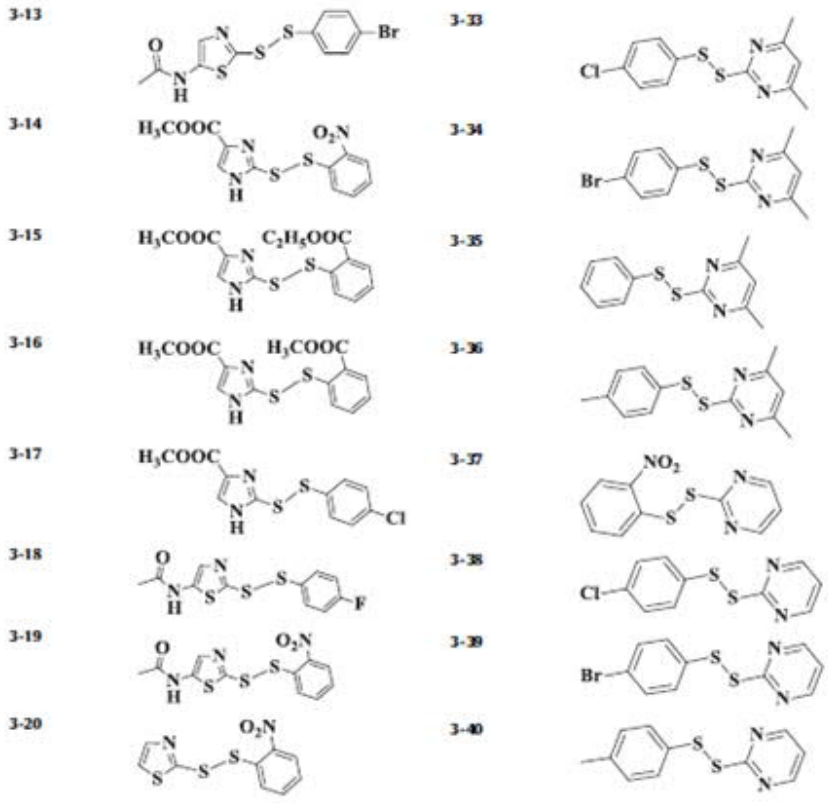

Figure 2(a)(ii). Disulphide inhibitors of SARS-CoV 3CL proteases taken from Wang [17]

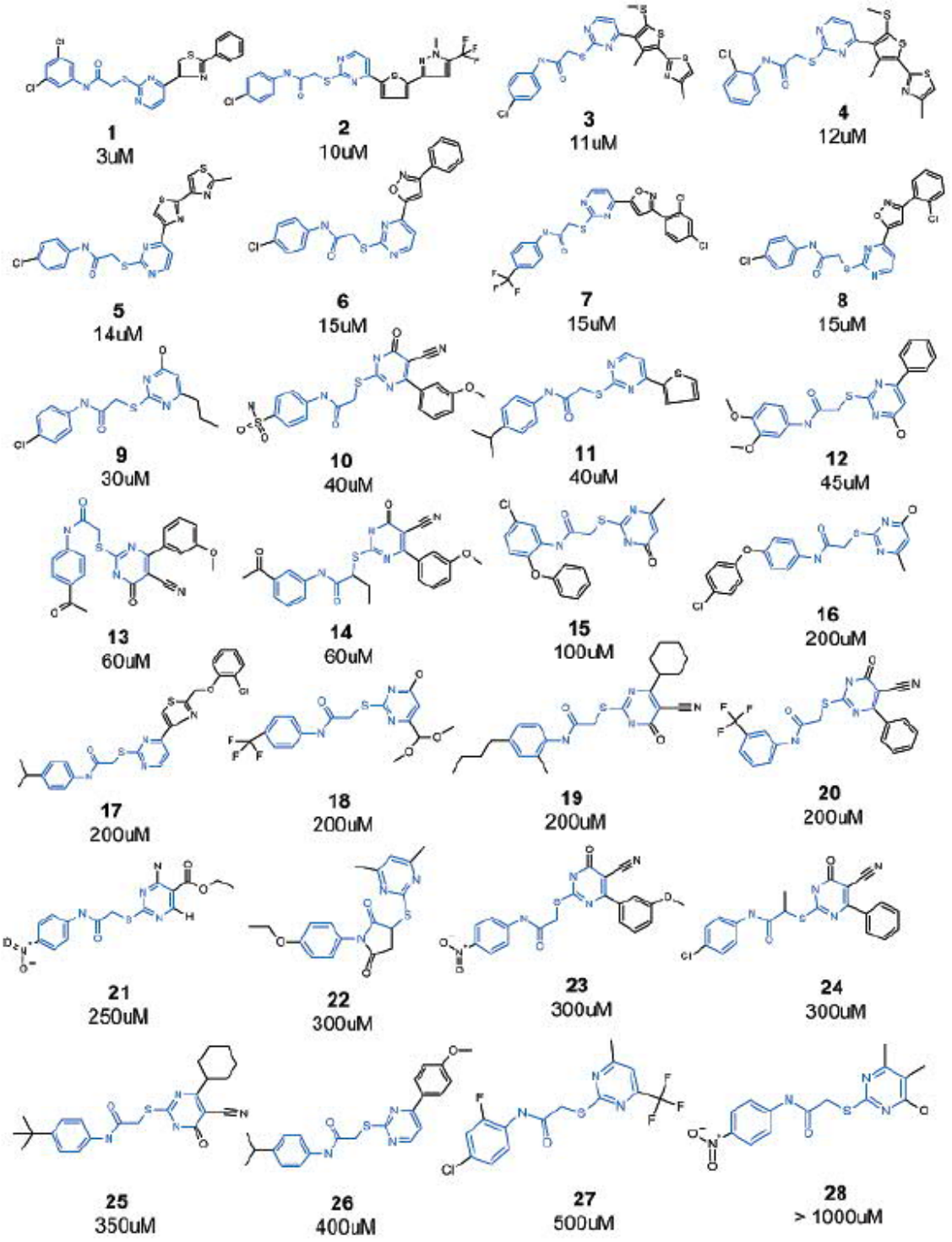

Figure 2(b). Inhibitors of SARS-CoV 3CL proteases taken from Tsai [18]

Table 1. Molecular specifiers for inhibitors of SARS-CoV 3C-like protease with IC_{50} values from Wang [17] and Tsai [18] used to derive equations 2 and 3.

Wang	IC_{50} μM	$\Delta G_{desolv,CDS}$ kcal/mol	$\Delta G_{lipo,CDS}$ kcal/mol	Dipole Moment D	Molec Volume cm^3/mol	HOMO eV	LUMO eV	HOMO-LUMO eV
3_1	1.871	-2.19	-5.82	3.79	156	-2.21	-6.66	-4.46
3_2	2.803	-3.94	-6.82	8.15	194	-2.23	-6.20	-3.97
3_3	3.675	-4.98	-7.69	9.08	190	-2.04	-6.60	-4.56
3_4	3.13	-5.6	-7.14	7.2	193	-2.71	-6.88	-4.17
3_5	1.506	-3.69	-6.32	3.57	162	-2.21	-6.44	-4.23
3_6	4.344	-5.03	-9.91	7.33	282	-1.71	-6.61	-4.89
3_7	4.1	-5.94	-9.19	8.53	264	-1.67	-6.22	-4.54
3_8	1.762	-4.76	-7.67	7.73	250	-2.73	-6.99	-4.26
3_9	5.654	-5.46	-9.32	4.15	334	-1.86	-6.64	-4.78
3_10	4.511	-5.62	-9.45	5.16	301	-1.91	-6.69	-4.77
3_12	2.626	-4.05	-6.72	8.05	258	-2.16	-6.10	-3.94
3_13	1.651	-3.26	-7.53	8.04	217	-2.17	-6.09	-3.93
3_14	2.075	-7.66	-6.32	15.19	208	-2.77	-6.59	-3.82
3_16	3.957	-7.48	-6.11	13.15	173	-2.06	-6.48	-4.42
3_17	4.126	-5.03	-6.7	8.92	190	-2.04	-6.61	-4.57
3_18	2.565	-4.55	-5.75	8.9	216	-2.16	-6.17	-4.01
3_19	1.947	-6.64	-6.47	13.93	245	-2.78	-6.27	-3.49
3_20	2.029	-4.96	-5.45	10.45	151	-2.80	-6.79	-3.99
3_21	1.25	-2.58	-5.65	6.37	163	-2.12	-6.55	-4.43
3_22	2.211	-2.78	-4.66	4.34	140	-2.15	-6.64	-4.49
3_24	2.555	-5.27	-5.9	10.44	175	-2.79	-6.60	-3.82
3_25	2.452	-5.22	-5.77	6.62	210	-2.21	-6.54	-4.33
3_26	1.679	-5.2	-3.61	10.33	188	-2.26	-6.81	-4.55
3_27	1.557	-5.11	-4.54	10.53	219	-2.22	-6.79	-4.57
3_28	1.713	-5.28	-3.76	11.73	156	-2.83	-7.04	-4.21
3_29	1.118	-5.5	-2.84	9.82	181	-2.31	-6.81	-4.50
3_31	0.516	-2.81	-3.26	5.05	146	-2.30	-6.96	-4.67
3_33	1.437	-2.33	-7.5	6	214	-1.85	-6.32	-4.47
3_34	1.121	-1.37	-8.13	5.94	200	-1.87	-6.54	-4.67
3_35	1.991	-2.52	-7.05	7.13	146	-1.70	-6.25	-4.55
3_36	1.495	-3	-7.29	7.53	179	-1.64	-6.20	-4.57
3_37	0.883	-4.56	-6.8	10.85	146	-2.83	-6.72	-3.89
3_38	0.684	-2.13	-7.15	4.74	182	-1.54	-6.64	-5.10
3_39	0.697	-1.35	-7.97	4.75	148	-1.55	-6.62	-5.07
3_40	1.522	-2.58	-7.08	5.59	162	-1.35	-6.48	-5.13
3_11*	5.794	-4.35	-8.09	3.58	270	-1.69	-6.23	-4.54
3_23*	3.321	-1.38	-6.58	3.92	212	-2.21	-6.64	-4.43
3_30*	1.264	-5.72	-4.86	7.05	185	-2.14	-6.44	-4.31

3_32*	0.921	-5.44	-7.5	11.28	211	-2.81	-6.62	-3.82
3_15*	5.954	-7.41	-7.08	13.15	189	-2.05	-6.46	-4.40
Tsai	**IC_{50}** μM	**$\Delta G_{desolv,CDS}$** kcal/mol	**$\Delta G_{lipo,CDS}$** kcal/mol	**Dipole Moment** D	**Molec Volume** cm^3/mol	**HOMO** eV	**LUMO** eV	**HOMO-LUMO** eV
1	3	-5.38	-11.65	12.61	335	-1.78	-6.05	4.27
1 alt**	3	-5.59	-11.75	9.22	320	-1.82	-6.12	4.30
2	10	-7.5	-9.58	12	317	-2.07	-5.84	3.76
3	11	-7.46	-11.82	11.45	334	-1.72	-5.48	3.77
4	12	-7.34	-11.85	7.12	343	-1.72	-5.48	3.77
5	14	-3.88	-9.68	13.32	281	-1.86	-6.05	4.19
6	15	-5.91	-10.71	12.55	327	-2.23	-6.10	3.87
7	15	-8.23	-10.03	5.84	320	-2.30	-6.33	4.03
8	15	-6.05	-11.12	13.9	366	-2.25	-6.09	3.84
9	30	-5.15	-8.57	11.42	278	-0.76	-6.06	5.30
10	40	-10.13	-8.43	7.47	268	-2.29	-6.26	3.97
11	40	-6.44	-9.52	10.48	279	-1.84	-5.88	4.04
12	45	-7.63	-9.29	8.15	280	-1.54	-5.89	4.35
13	60	-11.1	-9.76	10.18	195	-2.41	-6.26	3.85
14	60	-12.26	-10.1	4.25	258	-2.26	-6.19	3.93
15	100	-8.4	-10.42	11.84	292	-1.09	-5.99	4.90
16	200	-7.37	-9.62	16.24	253	-0.95	-3.51	2.56
17	200	-7.02	-12.01	10.04	371	-1.71	-5.92	4.21
18	200	-9.06	-5.48	12.68	235	-1.05	-6.27	5.22
19	200	-9.88	-10.67	13.7	305	-1.99	-5.95	3.96
20	200	-10.56	-7.72	7.59	289	-2.25	-6.46	4.22
21	250	-9.02	-8.31	13.5	288	-3.03	-6.39	3.36
22	300	-7.12	-9.26	4.9	241	-1.12	-6.01	4.89
23	300	-12.87	-8.7	8.83	309	-2.96	-6.25	3.29
24	300	-9.2	-9.47	7.94	264	-2.28	-6.30	4.02
25	350	-10.18	-9.57	12.81	331	-1.98	-5.91	3.93
26*	400	-8.03	-10.18	5.55	274	-1.56	-5.87	4.32
27*	500	-8.54	-8.96	9.63	299	-1.84	-6.16	4.32
28	>1000	-7.45	-7.7	17.19	227	-3.00	-6.12	3.12
Boceprevir		-13.13	-9.97	10.72	374	-2.13	-6.32	4.19
Telaprevir		-19.65	-12.53	8.25	464	-3.42	-5.13	1.71
13B α-ketoamide***		-13.37	-13.19	10.15	392	-2.41	-5.66	3.25

Footnotes: IC_{50} values from Wang [17] and Tsai [18], see Figure 2(a) for structures. Values* were outliers from derived equations 2 and 3. 1alt** was a different conformation from 1 showing only changes in dipole moment. See Figure 1 for structures of Boceprevir, Telaprevir, and13B α-ketoamide.

Table 2. Calculated inhibitory constants from Equations 1, 2(b) and 3(a) for a range of repurposed antivirals: eq 1 applies to the bat HKU4 3C-like protease, and eq 2(b) and 3(a) apply to the SARS-CoV 3C-like protease

	$\Delta G_{desolv,CDS}$ kcal/mol	$\Delta G_{lipo,CDS}$ kcal/mol	Dipole Mom D	Vol cm^3/mol	LUMO eV	HOMO eV	HOMO-LUMO eV	pIC_{50} Eq 1 µM	IC_{50} Eq 2(b) µM	IC_{50} Eq (a) µM
Chloroquine Neut	-3.4	-7.71	7.66	217	-1.28	-5.35	4.07	5.02	2.47	6.83
Chloroquine Ion	-6.19	-7.81	30.85	232	-1.31	-5.60	4.29	3.03	-0.10	252.66
Chloroquine Di-Ion	-7.98	-8.1	12.01	244	-2.22	-6.54	4.32	4.41	4.67	140.30
Hydroxychloro-quine Neut	-3.61	-7.93	7.97	261	-1.26	-5.45	4.18	4.97	2.64	12.24
Hydroxychloro-quine Ion	-6.3	-8.02	28.61	298	-1.29	-5.60	4.30	3.20	0.51	236.97
Hydroxychloro-quine Di-Ion	-8.09	-8.31	10.29	279	-2.20	-6.55	4.34	4.54	5.18	128.50
Favipiravir	-4.6	-1.03	8.96	82	-2.45	-6.46	4.01	4.81	0.75	120.69
Lopinavir	-13.93	-16.77	5.82	491	-0.08	-5.84	5.76	4.37	14.12	196.36
Remdesivir	-13.44	-10.79	12.76	381	-1.19	-5.97	4.78	4.01	10.12	292.61
Remdesivir TriPhosphate	-15.83	-6.99	18.29	267	-1.11	-5.97	4.85	3.41	9.97	462.01
GS441524	-5.39	-4.83	7.65	166	-1.17	-5.96	4.79	4.73	3.39	111.65
Saquinavir	-11.96	-14.84	10.08	501	-2.02	-5.92	3.89	4.54	9.99	145.83
Invermectin B1A	-18.95	-15.67	16.56	547	-0.73	-5.79	5.06	3.44	15.17	419.38
Ritonavir	-14.09	-14.74	8.81	644	-0.94	-6.16	5.22	4.23	12.58	229.96
Atazanavir	-18.24	-12.86	6.16	544	-1.26	-5.76	4.50	4.39	15.71	351.99
Nelfinavir	-11.77	-13.78	11.8	390	-0.93	-5.70	4.77	4.21	9.92	198.50
Ledipasvir	-17.12	-17.67	2.11	666	-1.73	-5.32	3.59	5.02	16.57	204.47
Velpatasvir	-14.99	-18.09	15.9	521	-1.45	-5.30	3.85	3.99	12.34	242.52
Nitazoxanide	-9.08	-6.79	13.18	172	-2.98	-6.83	3.85	4.36	4.49	180.25
Ruxolitinib	-4.35	-7.47	8.55	226	-1.26	-5.81	4.55	4.80	2.99	45.21
Baricitinib	-3.54	-7.06	6.86	259	-1.31	-5.84	4.53	4.97	2.57	14.32
Carfilzomib	-14.54	-15.94	5.38	580	-1.48	-5.94	4.46	4.67	13.60	185.52
Nafamostat Neut	-6.86	-9.19	15.39	296	-1.51	-5.37	3.86	4.32	3.82	134.72
Nafamostat Di-Ion	-9.51	-9.38	19.85	228	-2.13	-6.19	4.06	3.80	4.64	224.57
Ribavirin	-4.63	-5.52	9.85	174	-1.27	-7.23	5.95	4.34	2.45	91.93
Darunavir	-10.78	-10.68	8.72	331	-0.74	-5.75	5.01	4.40	9.12	199.62
Sofusbuvir	-14.37	-9.51	7.61	397	-1.32	-6.70	5.38	4.22	11.51	298.61
Galidesivir Neut	-2.55	-5.49	12.09	170	-0.47	-5.33	4.84	4.52	0.85	69.46
Galidesivir	-3.53	-5.62	14.51	179	-0.62	-5.86	5.24	4.19	1.06	109.40

Ion										
Dolutegravir	-8.9	-8.43	17.07	306	-1.75	-6.21	4.46	3.95	4.74	211.43
Efavirenz	-8.24	-5.37	10.3	199	-2.85	-4.68	1.82	5.08	4.15	159.97
Grazoprevir	-13.53	-13.61	12.7	521	-1.82	-5.94	4.12	4.20	10.51	234.73
Arbidol	-6.99	-10.45	10.84	263	-1.21	-5.34	4.13	4.62	5.36	94.98
Arbidol Ion	-9.4	-10.56	17.9	282	-1.38	-5.87	4.50	3.87	5.72	209.84
Imatinib	-2.4	-13.91	6	404	-1.87	-5.33	3.47	5.41	3.19	-144.50
Imatinib Ion	-5.47	-14.05	48.94	400	-1.88	-5.35	3.47	1.91	-3.19	251.37
Boceprevir	-13.13	-9.97	10.72	374	-2.13	-6.32	4.19	4.32	9.50	255.68
Telaprevir	-19.65	-12.53	8.25	464	-3.42	-5.13	1.71	4.79	14.95	354.60
13B α-ketoamide	-13.37	-13.19	10.15	392	-2.41	-5.66	3.25	4.60	10.43	202.15

Footnotes: Inhibitors colour coded in red are dominantly protonated at the physiological pH. Inhibitors colour coded in green are predicted to have high inhibitory capacity.